Advances in Intelligent Systems and Computing 1458

Jonathan Ansari · Sebastian Fuchs ·
Wolfgang Trutschnig ·
María Asunción Lubiano · María Ángeles Gil ·
Przemyslaw Grzegorzewski ·
Olgierd Hryniewicz *Editors*

Combining, Modelling and Analyzing Imprecision, Randomness and Dependence

Springer

Advances in Intelligent Systems and Computing 1458

Series Editor

Janusz Kacprzyk, *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Advisory Editors

Nikhil R. Pal, *Indian Statistical Institute, Kolkata, India*
Rafael Bello Perez, *Faculty of Mathematics, Physics and Computing, Universidad Central de Las Villas, Santa Clara, Cuba*
Emilio S. Corchado, *University of Salamanca, Salamanca, Spain*
Hani Hagras, *School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK*
László T. Kóczy, *Department of Automation, Széchenyi István University, Gyor, Hungary*
Vladik Kreinovich, *Department of Computer Science, University of Texas at El Paso, El Paso, USA*
Chin-Teng Lin, *Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan*
Jie Lu, *Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia*
Patricia Melin, *Graduate Program of Computer Science, Tijuana Institute of Technology, Tijuana, Mexico*
Nadia Nedjah, *Department of Electronics Engineering, University of Rio de Janeiro, Rio de Janeiro, Brazil*
Ngoc Thanh Nguyen, *Faculty of Computer Science and Management, Wrocław University of Technology, Wrocław, Poland*
Jun Wang, *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong*

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Jonathan Ansari · Sebastian Fuchs ·
Wolfgang Trutschnig · María Asunción Lubiano ·
María Ángeles Gil · Przemyslaw Grzegorzewski ·
Olgierd Hryniewicz
Editors

Combining, Modelling and Analyzing Imprecision, Randomness and Dependence

Springer

Editors
Jonathan Ansari
Department of Artificial Intelligence
and Human Interfaces
Paris Lodron Universität Salzburg
Salzburg, Austria

Sebastian Fuchs
Department of Artificial Intelligence
and Human Interfaces
Paris Lodron Universität Salzburg
Salzburg, Austria

Wolfgang Trutschnig
Department of Artificial Intelligence
and Human Interfaces
Paris Lodron Universität Salzburg
Salzburg, Austria

María Asunción Lubiano
Department of Statistics and Operational
Research and Mathematics Didactics
University of Oviedo
Oviedo, Spain

María Ángeles Gil
Department of Statistics and Operational
Research and Mathematics Didactics
University of Oviedo
Oviedo, Spain

Przemyslaw Grzegorzewski
Faculty of Mathematics and Information
Science
Warsaw University of Technology
Warsaw, Poland

Olgierd Hryniewicz
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-031-65992-8 ISBN 978-3-031-65993-5 (eBook)
https://doi.org/10.1007/978-3-031-65993-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

This volume contains peer-reviewed papers presented at the 11th International Conference on Soft Methods in Probability and Statistics (SMPS 2024), held in the beautiful city of Salzburg, Austria, from September 3rd to 6th, 2024. The series of biannual international conferences on Soft Methods in Probability and Statistics began in Warsaw in 2002 and has since successfully convened in Oviedo (2004), Bristol (2006), Toulouse (2008), Oviedo/Mieres (2010), Konstanz (2012), Warsaw (2014), Rome (2016), Compiègne (2018), and Valladolid (2022). SMPS 2024 continues this tradition, providing a forum for the latest advancements and discussions in the context of methodic developments in the fields of probability and statistics. This year's conference was organized by the Dependence Modeling working group of the department for Artificial Intelligence and Human Interfaces (AIHI), Faculty of Digital and Analytical Sciences, at the Paris Lodron Universität Salzburg.

The main objective of the SMPS conference series is to strengthen the dialogue between various research communities in order to cross-fertilize the fields and generate mutually beneficial activities. For SMPS 2024 we aimed at bringing together experts from diverse areas, fostering interactions between (soft) computing, probability theory, and statistics. These fields, originating from computer science and mathematics respectively, offer unique perspectives on data analysis. Soft computing provides quick, low-cost solutions inspired by nature, adaptable to different problems and models, making it successful in real-world applications. Its emphasis on understandable and interpretable solutions has made it popular also in economic fields. On the other hand, probability theory and statistics focus on creating robust, validated mathematical models, whose interpretation is mostly straightforward. It is our hope that the synergies of these fields and their interplay with modern techniques from machine learning will enhance intelligent data analysis methods, improving robustness to noise and applicability to larger datasets, while efficiently providing understandable solutions to real-world problems.

The SMPS 2024 covers a broad variety of topics ranging from classification, clustering, decision theory, dependence modeling, functional data analysis, optimal transport, robust statistics via core SMPS competencies like fuzzy sets, imprecise and generalized probabilities, random sets, and T-norms to machine learning and industrial applications. The approximately 70 papers in this volume were carefully selected through a peer-review process by international experts. The articles are organized in alphabetical order of the authors' names.

We are delighted that our keynote speakers, Siegfried Hörmann (Graz University of Technology, Austria), Susanne Saminger-Platz (Johannes Kepler University Linz, Austria), Steven Vanduffel (Vrije Universiteit Brussel, Belgium), Ingrid Van Keilegom (KU Leuven, Belgium), and Christian Wagner (University of Nottingham, UK), have accepted our invitation to present their thrilling latest research at the SMPS 2024.

No successful conference can be organized without the support of various people. We are very grateful to all contributing authors, invited speakers, program committee

members, and additional referees who made it possible to put together an attractive program for the conference. Special thanks go to the editor of the Springer series Advances in Soft Computing, Prof. Janusz Kacprzyk, and Springer-Verlag for their dedication to the production of this volume. We would also like to thank the University of Salzburg, IDA Lab Salzburg, and the AIHI department for their continuous support. Finally, our heartfelt thanks goes to María Asunción Lubiano for putting together the book of proceedings and to the local organizing committee for their outstanding effort in ensuring the smooth and enjoyable running of the conference.

Salzburg, Austria
May 2024

Jonathan Ansari
Sebastian Fuchs
Wolfgang Trutschnig
María Asunción Lubiano
María Ángeles Gil
Przemyslaw Grzegorzewski
Olgierd Hryniewicz

Organization

SMPS 2024 General Chairs

Jonathan Ansari	Paris-Lodron-University Salzburg, Austria
Sebastian Fuchs	Paris-Lodron-University Salzburg, Austria
Wolfgang Trutschnig	Paris-Lodron-University Salzburg, Austria

Executive Board (Core SMPS Group)

María Ángeles Gil	University of Oviedo, Spain
Przemyslaw Grzegorzewski	Polish Academy of Sciences, Warsaw, Poland
Olgierg Hryniewicz	Polish Academy of Sciences, Warsaw, Poland
María Asunción Lubiano	University of Oviedo, Spain

Program Commitee

Thomas Augustin	Ludwig-Maximilian-University Munich, Germany
Daniel Bartl	University of Vienna, Austria
Paula Brito	University of Porto, Portugal
Ana Colubi	King's College London, UK
Gert de Cooman	Ghent University, Belgium
Thierry Denoeux	Université de technologie de Compiègne, France
Nicolas Pascal Dietrich	Paris-Lodron-University Salzburg, Austria
Pierpaolo D'Urso	Sapienza University of Rome, Italy
Fabrizio Durante	University of Salento, Italy
Erol Egrioglu	Giresun University, Turkey
Alessio Farcomeni	University of Rome Tor Vergata, Italy
Maria Brigida Ferraro	Sapienza University of Rome, Italy
Peter Filzmoser	Vienna University of Technology, Austria
Alfonso García-Pérez	Universidad Nacional de Educación a Distancia, Spain
Paolo Giordani	Sapienza University of Rome, Italy
Francesca Greselin	University of Milano, Italy
Simon Hirlaender	Paris-Lodron-University Salzburg, Austria

Christoph Jansen	Ludwig-Maximilian-University Munich, Germany
Frank Klawonn	Helmholtz Centre for Infection Research, Germany
Vladik Kreinovich	University of Texas at El Paso, USA
Catherine Laflamme	Factory Planning and Production Management Fraunhofer Austria Research GmbH, Vienna, Austria
Patrick Benjamin Langthaler	Paris-Lodron-University Salzburg, Austria
María Asunción Lubiano	University of Oviedo, Spain
Enrique Miranda	University of Oviedo, Spain
Alfred Müller	University of Siegen, Germany
Marta Nai Ruscone	University of Genova, Italy
Ostap Okhrin	Dresden University of Technology, Germany
Elisa Perrone	Eindhoven University of Technology, Netherlands
Giovanni Puccetti	University of Milan, Italy
Una Radojičić	Vienna University of Technology, Austria
Dan Ralescu	University of Cincinnati, Ohio, USA
Marco Riani	University of Parma, Italy
Susanne Saminger-Platz	Johannes Kepler University Linz, Austria
Georg Schollmeyer	Ludwig-Maximilian-University Munich, Germany
Beatriz Sinova	University of Oviedo, Spain
Martin Štěpnička	University of Ostrava, Czech Republic
Pedro Terán	University of Oviedo, Spain
Valentin Todorov	UNIDO, Austria
Stefan Van Aelst	KU Leuven, Belgium
José Antonio Vilar	University of A Coruña, Spain
Johannes Wiesel	Carnegie Mellon University, USA
Sławomir Zadrożny	Polish Academy of Sciences, Warsaw, Poland
Georg Zimmermann	Paris-Lodron-University Salzburg, Austria

Local Organization Committee

Andrea Baumgartner	Paris-Lodron-University Salzburg, Austria
Lea Maislinger	Paris-Lodron-University Salzburg, Austria
Ulrike Ruprecht	Paris-Lodron-University Salzburg, Austria
Stefanie Steinmaßl	Paris-Lodron-University Salzburg, Austria

Contents

Aggregation of Random Initializations for Robust Linear Clustering 1
Pedro C. Álvarez-Esteban, Luis A. García-Escudero, and Agustín Mayo-Iscar

An Empirical Study on New Model-Free Multi-output Variable Selection Methods 9
Jonathan Ansari, Eva Lütkebohmert, and Marcus Rockel

Minimizing the Expected Absolute Difference with a Fixed Continuous Copula 18
Juan Baz, Irene Daz, and Susana Montes

An EM Stopping Rule for Avoiding Degeneracy in Gaussian-Based Clustering with Missing Data 27
Christophe Biernacki and Vincent Vandewalle

Limit Theorems for Convex Expectations 36
Jonas Blessing and Michael Kupper

The Joint Weighted Average (JWA) Operator 44
Stephen B. Broomell and Christian Wagner

A Bayesian Modeling Approach to Fuzzy Data Analysis 59
Antonio Calcagnì and Przemyslaw Grzegorzewski

Compound Conditionals and Fuzzy Sets 67
Lydia Castronovo and Giuseppe Sanfilippo

Sharp Polynomial Upper Bound on the Variance 76
Marco de Angelis

Hierarchical Clustering and CoClust Algorithm: A Nested Procedure to Analyse Sustainable Heating Data 85
F. Marta L. Di Lascio and Roberta Pappadà

Fuzzy Clustering Implementations for Big Data in R 93
Vincenzo Di Perna and Maria Brigida Ferraro

On Topologically Typical Bivariate Extreme Value Copulas 102
Nicolas Pascal Dietrich and Wolfgang Trutschnig

Semi-supervised Learning Guided by the Generalized Bayes Rule Under Soft Revision 110
Stefan Dietrich, Julian Rodemann, and Christoph Jansen

Dissimilarity-Based Clustering with Soft Proximity Constraints 118
Fabrizio Durante and Roberta Pappadà

Copula-Based Fuzzy Clustering of Count Data with Total Variation Distance 126
Pierpaolo D'Urso, Livia De Giovanni, Lorenzo Federico, and Vincenzina Vitale

Missing in Survival Analysis 134
Sam Efromovich and Lirit Fuksman

Causal Markov Categories and Possibility Theory 142
Tobias Fritz and Pedro Terán

Sensitivity Analysis on the Choice of the Metric on Cronbach's α Coefficient for Interval-Valued Data in Questionnaires 149
José García-García, María Ángeles Gil, and María Asunción Lubiano

On a Linear Regression Model for Star-Shaped Set-Valued Random Elements 158
José Graña-Colubi, Gil González-Rodríguez, and Ana Belén Ramos-Guajardo

Two-Sample Depth-Based Test for Dispersion in Fuzzy Environment 166
Przemyslaw Grzegorzewski and Anna Kozak

Deep Meta Reinforcement Learning for Rapid Adaptation In Linear Markov Decision Processes: Applications to CERN's AWAKE Project 175
Simon Hirlaender, Sabrina Pochaba, Lamminger Lukas, Andrea Santamaria Garcia, Chenran Xu, Jan Kaiser, Annika Eichler, and Verena Kain

Aggregation of Inhomogeneous Non-stationary Time Series Using P-Boxes 184
Olgierd Hryniewicz and Katarzyna Kaczmarek-Majer

Entropy-Based Search for the Most Informative Belief Functions 192
Radim Jiroušek and Václav Kratochvíl

Estimation of Conditional Value-at-Risk in Linear Model 200
Jana Jurečková, Jan Picek, and Jan Kalina

Common Models of Errors in Variables 208
Henrik Kaiser

Data Augmentation with ChatGPT for Assessing Subject Alignment 217
Louisa Kontoghiorghes and Ana Colubi

The Impact of Sentiment in Social Network Communication 225
Soveatin Kuntur and Anna Cena

Constructing Measures of Dependence Via Sensitivity of Conditional Distributions 234
Patrick B. Langthaler, Jonathan Ansari, Sebastian Fuchs, and Wolfgang Trutschnig

Smoke Emission Detection by Deep Learning 241
Jo-Huan Lee and Chi-Bin Cheng

Quantifying Directed Dependence with Kendall's Tau 249
Carsten Limbach and Sebastian Fuchs

Some Notes on Bivariate Lower Semilinear Copulas 256
Lea Maislinger and Wolfgang Trutschnig

Addressing Incomplete Data in Two-Way Contingency Tables with Three-Level Variables 264
Pitchayanin Makapawee, Kanyawee Kamkongkaew, and Monchai Kooakachai

Exploring Deep Reinforcement Learning Algorithms for Enhanced HVAC Control 273
Antonio Manjavacas, Alejandro Campoy-Nieves, Miguel Molina-Solana, and Juan Gómez-Romero

Clustering on the d-Dimensional Sphere: Latest Trends and the Role of Poisson Kernel-Based Density Models 281
Marianthi Markatou

Aggregation of the Distortion Models Induced by the KL Divergence and Euclidean Distance 286
Ignacio Montes

Regularised Optimal Self-transport Is Approximate Gaussian Mixture Maximum Likelihood 294
Gilles Mordant

Stochastic Orders Under Uncertainty 302
Alfred Müller

Conditions for the Multivariate Stochastic Order Under Dependence Uncertainty 309
Alfred Müller and Marco Scarsini

Variable Selection for Clustering Three-Way Data 317
Mackenzie R. Neal and Paul D. McNicholas

Functional Outlier Detection 325
Jeremy Oguamalam, Una Radojičić, and Peter Filzmoser

Patchwork Methods for Multivariate Quasi-Copulas 334
Matjaž Omladič, Tomaž Košir, Damjana Kokol Bukovšek, Blaž Mojškerc, and Nik Stopar

Optimal Change-Point Sequential Detection in Autoregressive Time Series 342
Serguei Pergamenchtchikov and Roman Tenzin

Multi-way Contingency Tables with Uniform Margins 349
Elisa Perrone, Roberto Fontana, and Fabio Rapallo

Low-Rank Approximation of Data Matrices Using Robust Sparse Principal Component Analysis 357
Pia Pfeiffer and Peter Filzmoser

Multi-agent Reinforcement Learning and Its Application to Wireless Network Communication 363
Sabrina Pochaba, Peter Dorfinger, Matthias Herlich, Roland Kwitt, and Simon Hirlaender

Order Determination in Second-Order Source Separation Models Using Data Augmentation 371
Una Radojičić and Klaus Nordhausen

Clustering Ordinal Data Via Parsimonious Models 380
Monia Ranalli and Roberto Rocci

Robustness in LiNGAM 388
Jakob Raymaekers

Comparing Bivariate Random Vectors by Means of Statistical Preference 396
Julián Ros, Raúl Pérez-Fernández, and Ignacio Montes

New Decision Rules for Fuzzy Statistical Inferences 405
Julien Rosset and Laurent Donzé

Learning in Reproducing Kernel Hilbert Spaces for Orbits of Iterated Function Systems 413
Priyanka Roy and Susanne Saminger-Platz

A Comparative Analysis of Aggregation Rules for Coherent Lower Previsions 421
Juan J. Salamanca, Ignacio Montes, and Enrique Miranda

Copulas and Triangular Norms: Selected Commonalities and Differences 429
Susanne Saminger-Platz

Characterizations of Multivariate Implicit Dependence Copulas 440
Tippawan Santiwipanont, Songkiat Sumetkijakan, and Noppawit Yanpaisan

Python-Based Reinforcement Learning on Simulink Models 449
Georg Schäfer, Max Schirl, Jakob Rehrl, Stefan Huber, and Simon Hirlaender

Applied Statistics in Industry: Defining an Appropriate Target Variable and Analysing Factors Affecting Aluminium Ingot Quality 457
Manuela Schreyer, Marco Tschimpke, Alexander Gerber, Steffen Neubert, and Wolfgang Trutschnig

Fuzzy Medoids Based on an L^1-Type Distance 466
Beatriz Sinova, Sergio Palacio Vega, and María Ángeles Gil

Fuzzy S-Estimators 474
Beatriz Sinova and Stefan Van Aelst

Reversing Inhomogeneous and Imprecise Markov Chains 483
Damjan Škulj

Minimum Intrinsic Dimension Scaling for Entropic Optimal Transport 491
Austin J. Stromme

Creation of Probabilistic Models by Experts: A Case Study 500
Radek Švadlenka and Radim Jiroušek

Possibilistic Variables and Akian Semirings 509
Pedro Terán

Robust Sparse PCA with R 518
Valentin Todorov

Dependent Censoring Based on Copulas 526
Ingrid Van Keilegom

Wavelet Variance Based Robust Estimation of Composite Stochastic Models 532
Maria-Pia Victoria-Feser, Stephane Guerrier, and Roberto C. Molinari

Intersections of Randomly Translated Sets 540
Tommaso Visonà

Hierarchical Variable Clustering Based on Measures of Predictability 548
Yuping Wang and Sebastian Fuchs

Large Deviation Theory Without Probability Measures 554
José M. Zapata

Author Index 563

Aggregation of Random Initializations for Robust Linear Clustering

Pedro C. Álvarez-Esteban, Luis A. García-Escudero(✉), and Agustín Mayo-Iscar

Departamento de Estadística e I.O. and IMUVA, Universidad de Valladolid, Paseo de Belén S/n, 47011 Valladolid, Spain
{pedrocesar.alvarez,lagarcia,agustin.mayo.iscar}@uva.es

Abstract. Outliers can be very harmful to clustering methods, to the extent that even a small fraction of outliers can significantly impact their performance. Consequently, the use of robust clustering procedures has been advocated. A robust linear clustering procedure will be reviewed here, where robustness is pursued by applying a trimming approach. The proposed algorithm for its practical implementation is based on concentration steps, analogous to those applied in well-known high-breakdown robust multivariate procedures. However, correct initialization is key for the algorithm's adequate performance. A large number of random initializations is often required to guarantee avoidance of local minima of the robust linear clustering target function. An alternative "ensemble initialization" method will be given to try to overcome this problem.

Keywords: Clustering · Robustness · Outliers · Affine subspaces · Initialization · Ensembles

1 Robust Linear Clustering Based on Trimming

Modifying a small proportion of observations can severely impact the results of clustering methods, and their outcomes may be negatively affected by anomalous or outlying measurements [3]. Accordingly, several robust clustering approaches have been proposed to mitigate the influence of outliers on clustering results [5,10].

Here, we focus on a trimming-based approach for robust linear clustering. Trimming methods in statistics involve discarding a fraction α of observations, ideally those that are the most outlying, before applying standard statistical tools. The reliance on standard clustering methods after trimming makes this approach easy to understand and explain.

To be more precise, we consider an impartial trimming approach, where "impartial" implies that the dataset itself determines which observations to be trimmed. This impartial trimming approach is also fundamental to well-known robust procedures such as LTS (Least Trimmed Squares), LMS (Least Median

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 1–8, 2024.
https://doi.org/10.1007/978-3-031-65993-5_1

of Squares), MCD (Minimum Covariance Determinant), and MVE (Minimum Volume Ellipsoid) (see [9]).

Given a sample $\{x_1, ..., x_n\} \subset \mathbb{R}^p$, robust clustering methods based on impartial trimming produce partitions of the n observations into K clusters, while leaving a fraction $\alpha \in [0, 1)$ of observations trimmed. These trimmed observations are those that least conform to the assumed global clustering structure. The first robust clustering approach based on impartial trimming was introduced in [1], where *trimmed K-means* was proposed as a robust extension of the classical K-means algorithm. The trimmed K-means method seeks K centres $m_1, ..., m_K$ and an index partition $\{R_0, R_1, ..., R_K\}$ of $\{1, 2, ..., n\}$ such that $\#R_0 = [n\alpha]$ and minimizes

$$\sum_{k=1}^{K} \sum_{i \in R_k} \|x_i - m_k\|^2. \tag{1}$$

Since the sum in (1) ranges from 1 to K, a proportion α of the observations in R_0 are not considered in the target function. The proportion α of observations in the optimal R_0 set are the trimmed ones.

Starting from the trimmed K-means, it is reasonable to assume that clusters are concentrated around lower-dimensional affine subspaces as the dimension p increases, thus taking into account dependence structures within clusters. In fact, this type of dimensionality reduction is analogous to that employed when assuming a single $K = 1$ homogeneous population with no outliers, and classical Principal Components Analysis (PCA) is applied. Its robust extension to robust clustering based on trimming would be the *robust linear grouping algorithm* (RLG) introduced in [6]. The RLG searches for K affine subspaces $h_1, ..., h_K$ with "intrinsic" dimensions $q_1, ..., q_K$ $(q_k < p)$ and a partition $\{R_0, R_1, ..., R_K\}$ of $\{1, 2, ..., n\}$ such that $\#R_0 = [n\alpha]$ and minimizes

$$\sum_{k=1}^{K} \sum_{i \in R_k} \|x_i - \mathrm{Pr}_{h_k}(x_i)\|^2, \tag{2}$$

where $\mathrm{Pr}_h(x)$ denotes the orthogonal projection of $x \in \mathbb{R}^p$ onto the affine subspace h. The original formulation in [6] assumed $q_1 = ... = q_K = p - 1$, but a more general statement assuming unequal intrinsic dimensions is straightforward. Note also that RLG reduces to the LTS-PCA approach in [7] in the $K = 1$ case.

Figure 1 shows the result of applying RLG to a simulated data set in $\mathbb{R}^3$ clustered around $K = 3$ affine subspaces, where one of them has an intrinsic dimension of 1 (a single line), and the others two have intrinsic dimensions of 2 (two planes). Additionally, a 6% proportion of scattered outliers has been added. RLG has been applied with $K = 3$, $\alpha = 0.06$, $q_1 = 1$ and $q_2 = q_3 = 2$.

The implementation of RLG (and other robust clustering methods based on impartial trimming) involves solving complex combinatorial problems because, in principle, all possible partitions of our n observations into $K + 1$ subsets are considered (with one of the subsets containing a fixed fraction α of observations to be trimmed). Fortunately, "concentration steps" or C-step introduced in [11],

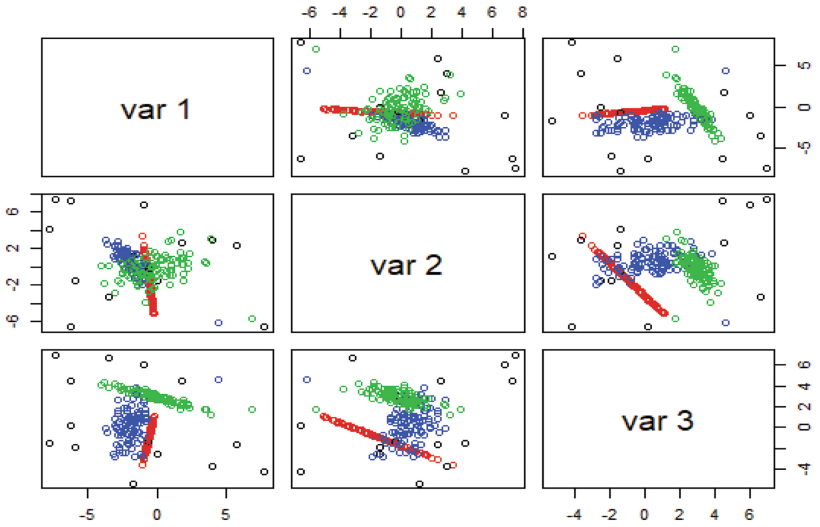


Fig. 1. RLG results with $K = 3$ clusters (in red, green and blue), $q_1 = 1$, $q_2 = q_3 = 2$, and a trimming level $\alpha = 0.06$ for a simulated data set. Trimmed observations are indicated by the symbol "○".

designed to compute the MCD estimator, can be adapted to this clustering framework and successfully applied. A description of how C-steps can be applied in the case of the RLG will be provided in Sect. 2.

2 Algorithm and Difficulties in Initialization

The algorithm proposed for implementing RLG algorithm in [6] can be viewed as an extension of the algorithm for the trimmed k-means in [4]. However, in its C-steps, distances to centers are replaced by orthogonal distances to affine subspaces. Moreover, instead of simple means to update K centers, the application of PCA is necessary to update the K approximating affine subspaces.

With this type of C-steps, it is easy to observe that the target function (2) monotonically decreases at each iteration,thereby leading to a local minimum. However, several random initializations are necessary to increase the likelihood of finding a close approximation to a global minimum. In fact, the high non-convexity of the target functions considered makes the use of multiple initializations absolutely essential, as is the case with other high-breakdown robust procedures such as the LTS and the MCD estimators, or even when applying the simpler (untrimmed) k-means method.

The random initializations for this algorithm can be obtained from randomly selecting subsets of observations from the available data set. One straightforward approach is to select $\sum_{k=1}^{K}(q_k + 1)$ observations, which is the minimum number of observations needed (in general position) to initialize K affine subspaces with intrinsic dimension q_1,...,q_K.

In describing the algorithm, we will use the notation `nstart` to denote the number of random initializations and `csteps` to represent the number of C-steps applied.

1. Consider `nstart` random initializations of the initial affine subspaces $h_1^0, ..., h_K^0$ with intrinsic dimensions $q_1, ..., q_K$, respectively, and set $l = 1$ in each of these initializations.
   1.1 Keep the proportion $1-\alpha$ of the closest observations, in term of orthogonal distances, to the affine subspaces $\{h_k^{l-1}\}_{k=1}^K$ in the previous iteration. If
   $$d_i = \min_{k=1,...,K} \|x_i - \mathrm{Pr}_{h_k^{l-1}}(x_i)\|^2,$$
   we sort these values as $d_{(1)} \leq d_{(2)} \leq ... \leq d_{(n)}$ and take
   $$H = \{i : d_i \leq d_{(n-[n\alpha])}\},$$
   and $H = H_1 \cup ... \cup H_K$, such that
   $$H_k = \{i \in H \text{ such that } d_i = \|x_i - \mathrm{Pr}_{h_k^{l-1}}(x_i)\|^2.$$
   1.2 The previous affine subspaces are updated in $\{h_k^l\}_{k=1}^K$ by solving K PCA problems, where the best approximating affine subspace of intrinsic dimension q_k is considered for the observations with indices in H_k, $k = 1, ..., K$.
   1.3 Steps 1.1 and 1.2 are alternated `csteps` times ($l = 1, ...,$ `csteps`) or until a stop rule is met.
2. The affine subspaces associated with the minimum value found for the target function (2) are returned as the output of the algorithm.

As mentioned, we are dealing with complex combinatorial optimization problems, and the effectiveness of the proposed algorithm heavily relies on having an efficient initialization among the `nstart` random initializations. The initialization based on selecting $\sum_{k=1}^K (q_k + 1)$ random observations ideally requires choosing them free of outliers, but also necessitates a correct partition of them for initializing the K affine subspaces. However, ensuring this correct initialization is challenging unless an extremely large number of random initializations (`nstart`) are attempted. This renders the algorithm infeasible when the number of clusters K and/or their intrinsic dimensions q_k are not very small.

3 Ensemble Initialization

The final output of the algorithm presented in Sect. 2 is the RLG partition corresponding to the "best" iterated initialization, i.e. the result of the C-steps iterated solution which has led to the smallest value in (2). However, this approach implies that all the partially correct information from the C-steps in the other `nstart`-1 random initialization is completely neglected.

"Ensemble clustering" approaches are popular in clustering (see, e.g., [2,12] or [8]). The main principle behind them is that the combination of information from different clustering partitions can improve our final clustering result. This same principle can be extended to the initialization of the RLG algorithm. The idea here is that the combination (ensemble) from all the random initializations after C-steps could be more informative than the result of the previously mentioned "best" iterated single initialization. This results in an *ensemble initialization* which aims to take advantage of the partially correct information of all the `nstart` iterated random initializations.

The ensemble initialization approach can be described as follows:

1. Consider `nstart` random initializations with `cstep` concentration steps to get $\{\mathcal{C}_b\}_{b=1}^{\texttt{nstart}}$ partitions $\mathcal{C}_b = \{R_0^b, R_1^b, ..., R_K^b\}$ of the indices $\{1, 2, ..., n\}$, with R_0^b denoting the indices of the trimmed observations in the b-th random initialization, $b = 1, ..., \texttt{nstart}$.
2. Compute a symmetric affinity matrix A with n rows and n columns and with terms
$$A_{ii'} = \frac{1}{\texttt{nstart}} \#\big\{b \text{ such that } \{i, i'\} \subset R_k^b \text{ for any } R_k^b \text{ with } k \neq 0\big\}.$$
This means that $A_{ii'}$ represents the proportion of random initializations where the pair of observations x_i and $x_{i'}$ are co-clustered, and neither of them are trimmed.
3. Compute
$$A_i = \sum_{i'=1}^{n} A_{ii'}, \tag{3}$$
for $i = 1, ..., n$, and sort them as $A_{(1)} \leq A_{(2)} \leq ... \leq A_{(n)}$. The initially non-trimmed observations are those with indices in $H = \{i : A_{(i)} \geq A_{([\alpha n]+1)}\}$, and the trimmed ones are those with indices in $H_0 = \{1, 2, ..., n\} \setminus H$.
4. A hierarchical clustering algorithm (or, alternatively, a spectral clustering approach) is applied to the observations $\{x_i : i \in H\}$ by using a dissimilarity matrix given by $d_{ii'} = 1 - A_{ii'}$. This way, we obtain a partition of H as $H = H_1 \cup ... \cup H_K$. The partition $H_0 \cup H_1 \cup ... \cup H_K$ will be our ensemble initialization.
5. The ensemble initialization $H_0 \cup H_1 \cup ... \cup H_K$ can be further improved by applying additional C-steps.

We would encounter higher affinities values $A_{ii'}$ for pairs of observations x_i and $x_{i'}$ that, in more iterated random initializations, end up in the same cluster. Alternatively, outlying observations are frequently trimmed and thus are not co-clustered with non-outlying ones. This causes outlying observations to exhibit small affinity values and, consequently, small A_i values in (3) as well.

4 An Illustrative Example

To illustrate the interest of the proposed approach, we consider a simple example. We generate $n = 100$ observations in $\mathbb{R}^{20}$ grouped around $K = 3$ affine subspaces

of dimensions $q_1 = q_2 = q_3 = 5$ (with 33, 33 and 34 observations in each). With this aim, three subsets, each comprising q_j orthonormal vectors in $\mathbb{R}^{20}$, are randomly selected. The 100 observations in the dataset are obtained by multiplying these vectors with weights uniformly distributed in the interval $[0, 1]$, and by adding normally distributed isotropic random noise in $\mathbb{R}^{20}$. Later, 6 of these observations are replaced by noisy observations, where their coordinates are randomly generated uniformly in an hypercube that includes the initial 100 observations.

The RLG algorithm in Sect. 2 is first applied with the correct values of the parameters, i.e. $K = 3$, $q_1 = q_2 = q_3 = 5$ and $\alpha = 0.06$, considering `nstart=200` and `csteps=10`. The partition associated with the best iterated random initialization is shown in Table 1 (left). We observe that this solution is clearly deficient because the three original clusters are not adequately detected, although 5 out of the 6 added outliers are correctly trimmed.

Table 1. Rows represent the true labels (with 0 indicating outliers), and columns represent the cluster assignments (with 0 indicating trimmed observations). The table on the left displays the best single iterated solution, while the table on the right displays the solution from the ensemble initialization.

	Best init.				Ensemble		
	0	1	2	3	0	1	2	3
0	5	0	0	1	6	0	0	0
1	0	9	11	7	0	30	0	0
2	1	0	12	24	0	0	32	0
3	0	21	9	0	0	0	0	32

The partially correct information from all the `nstart=100` iterated random initialization is aggregated in the affinity matrix A, shown in the heatmap in Fig. 2. Near the diagonal, we observe a structure of 3 boxes associated with higher affinities corresponding to the three generated clusters. Additionally, we can see 6 rows and 6 columns displaying small affinities, which exactly correspond to the 6 added outliers. After removing these 6 rows and 6 columns with the smallest values of A_i, we apply Ward's hierarchical clustering procedure. By cutting the associated dendrogram into $K = 3$ clusters to obtain the ensemble initialization, 10 additional C-steps are applied starting from that ensemble initialization. The results obtained are presented in Table 1 (right).

We observe a notable improvement in the clustering results with the ensemble initialization, where the three clusters and the added contamination are perfectly recovered. The additional computing time to compute the affinity matrix, apply hierarchical clustering and perform the final C-steps is not very high compared to the total time required for performing the C-steps for the `nstart=200` random initializations.

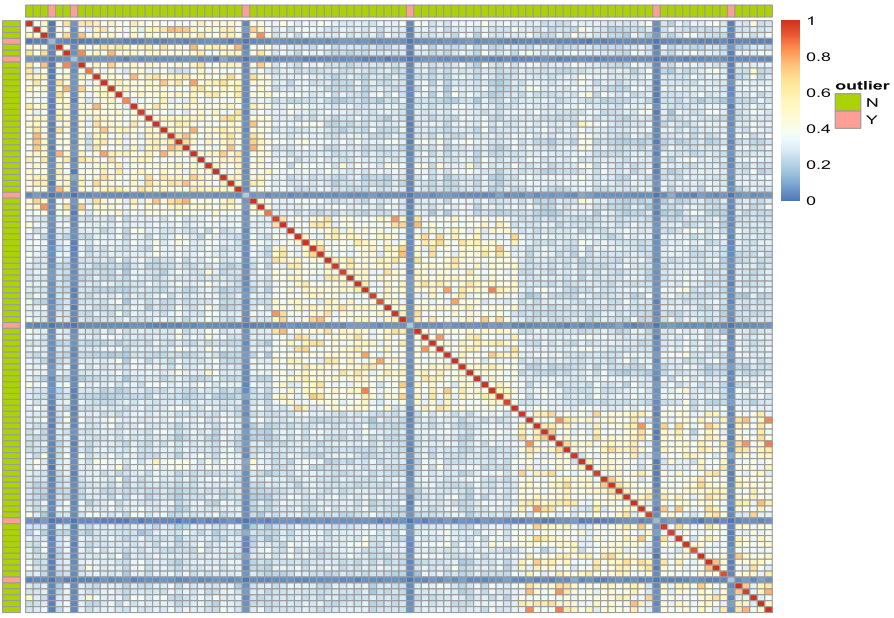


Fig. 2. Heatmap of the affinity matrix A obtained from the `nstart=100` random initializations. The position of the added outliers are indicated in the margins of this heatmap.

Acknowledgements. Research partially supported by grant PID2021-128314NB-I00 funded by MCIN/AEI/ 10.13039/501100011033/FEDER.

References

1. Cuesta-Albertos, J.A., Gordaliza, A., Matrán, C.: Trimmed k-means: an attempt to robustify quantizers. Ann. Stat. **25**, 553–576 (1997)
2. Fred, A.L.N., Jain, A.K.: Combining multiple clusterings using evidence accumulation: IIEEE Trans. Pattern Anal. Mach. Intell. **6**, 835–850 (2005)
3. García-Escudero, L.A., Gordaliza, A.: Robustness properties of k-means and trimmed k-means. IIEEE Trans. Pattern Anal. Mach. Intell. **94**, 956–969 (1999)
4. García-Escudero, L.A., Gordaliza, A., Matrán, C.: Trimming tools in exploratory data analysis. J. Comput. Graph. Stat. **12**, 434–449 (2003)
5. García-Escudero, L.A., Gordaliza, A., Matrán, C., Mayo-Iscar, A.: A review of robust clustering methods. Adv. Data Anal. Classif. **4**, 89–109 (2010)
6. García-Escudero, L.A., Gordaliza, A., San Martín, R., Van Aelst, S., Zamar, R.: Robust linear clustering. J. R. Stat. Soc. Series B Stat. Methodol. 71, 301–318 (2009)
7. Maronna, R.: Principal components and orthogonal regression based on robust scales. Biometrics **47**, 264–273 (2005)
8. Lipor, J., Hong, D., Tan, Y.S., Balzano, L.: Subspace clustering using ensembles of k-subspaces. Inf. Inference **10**, 73–107 (2021)

9. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley-Interscience, New York (1987)
10. Ritter, G.: Robust Cluster Analysis and Variable Selection. CRC Press, Boca Raton (2014)
11. Rousseeuw, P.J., van Driessen, K.: A fast algorithm for the minimum covariance determinant estimator. Technometrics **41**, 73–107 (1999)
12. Strehl, A., Ghosh, J.: Cluster ensembles—a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3**, 583–617 (2002)

An Empirical Study on New Model-Free Multi-output Variable Selection Methods

Jonathan Ansari[1], Eva Lütkebohmert[2], and Marcus Rockel[2(✉)]

[1] Paris Lodron Universität Salzburg, Salzburg, Austria
jonathan.ansari@plus.ac.at
[2] Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
eva.luetkebohmert@finance.uni-freiburg.de,
marcus.rockel@finance.uni-freiburg.de

Abstract. Chatterjee's rank correlation has attracted a lot of attention since it is able to characterize independence between a random variable Y and a random vector $\mathbf{X}$ as well as perfect directed dependence of Y on $\mathbf{X}$. Further, it satisfies an information gain inequality and also characterizes conditional independence, which—thanks to a fast estimator—allows a model-free variable selection in almost linear time. In this paper, we compare two multi-output variable selection methods based on recently investigated extensions of Chatterjee's rank correlation to a multivariate output vector $\mathbf{Y}$. In an empirical study, we investigate the relevance of the five Fama French factors to explain returns of six big tech companies over the time period from 2013 to 2024.

Keywords: Chatterjee's rank correlation · Conditional independence · Dependence measure · Feature selection

1 Preliminaries

Commonly known correlation measures between random variables X and Y such as the Pearson correlation, Spearman's rho or Kendall's tau have the property that they attain the value 0 if X and Y are independent and the value $+1$ or -1 if and only if X and Y are perfectly positively or negatively (linear) dependent. However, the value 0 does not necessarily imply that X and Y are independent. For instance, $Y = X^2$ perfectly depends on X even though the correlation is 0 for a standard normal X. In the last decade, an increasing number of scientific papers have been published on measures that overcome these undesirable properties. Particularly interesting is the class of *dependence measures* $\kappa(\mathbf{Y}|\mathbf{X})$ (a.k.a. measures of regression dependence or measures of predictability) for random vectors $\mathbf{X}$ and $\mathbf{Y}$, which satisfy the axioms:

(I) $\kappa(\mathbf{Y}|\mathbf{X}) \in [0,1]$,
(II) $\kappa(\mathbf{Y}|\mathbf{X}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent, and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 9–17, 2024.
https://doi.org/10.1007/978-3-031-65993-5_2

(III) $\kappa(\mathbf{Y}|\mathbf{X}) = 1$ if and only if there exists a measurable (not necessarily monotone) function $\boldsymbol{f}\colon \mathbb{R}^p \to \mathbb{R}^q$ such that $\mathbf{Y} = \boldsymbol{f}(\mathbf{X})$ almost surely, where p is the dimension of $\mathbf{X}$ and q the dimension of $\mathbf{Y}$.

A particular such functional is *Chatterjee's rank correlation* which is defined for a real-valued, non-degenerate random variable Y and a random vector $\mathbf{X}$ by

$$\xi(Y|\mathbf{X}) := \frac{\int_{\mathbb{R}} \mathrm{Var}(P(Y \geq y \mid \mathbf{X}))\, \mathrm{d}P^Y(y)}{\int_{\mathbb{R}} \mathrm{Var}(1_{\{Y \geq y\}})\, \mathrm{d}P^Y(y)}, \tag{1}$$

see [2,3]. Some further very useful properties of ξ are that it satisfies the information gain inequality $\xi(Y|\mathbf{X}) \leq \xi(Y|(\mathbf{X}, \mathbf{Z}))$ for all Y, $\mathbf{X}$, and $\mathbf{Z}$ and that it characterizes conditional independence, i.e., $\xi(Y|\mathbf{X}) = \xi(Y|(\mathbf{X}, \mathbf{Z}))$ if and only if Y and $\mathbf{Z}$ are conditionally independent given $\mathbf{X}$. Moreover, it admits a strongly consistent and asymptotically normal estimator given by

$$\xi_n(Y|\mathbf{X}) = \frac{\sum_{k=1}^n (n \min\{R_k, R_{N(k)}\} - L_k^2)}{\sum_{k=1}^n L_k(n - L_k)}, \tag{2}$$

where, for a random sample $\{\mathbf{X}_k, Y_k\}_{1 \leq k \leq n}$ from $(\mathbf{X}, Y)$, R_k denotes the rank of Y_k among $Y_1, \ldots, Y_n$, i.e., the number of j such that $Y_j \leq Y_k$, and L_k denotes the number of j such that $Y_j \geq Y_k$. For each k, the number $N(k)$ denotes the index j such that $\mathbf{X}_j$ is the nearest neighbour of $\mathbf{X}_k$ with respect to the Euclidean distance, where ties are broken uniformly at random. All these appealing properties allow a fast, model-free variable selection method established in [2], noting that ξ_n can be computed in $O(n \log n)$ time.

2 Multi-output Dependence Measures

While various variable selection methods are known in the literature for a single output variable, methods for multi-output variable selection are, apart from linear models, generally challenging and have been rarely investigated. The above described approach via dependence measures can be extended by generalizing Chatterjee's rank correlation to multiple output variables. There are two recently studied extensions—the first is a direct extension of Chatterjee's rank correlation using a conditioning argument.

Definition 1. Multidimensional Chatterjee rank correlation; [1]). *Let* $\mathbf{Y}$ *be a* q*-dimensional random vector with non-degenerate components. Then define*

$$T(\mathbf{Y}|\mathbf{X}) := 1 - \frac{q - \sum_{i=1}^q \xi(Y_i|(\mathbf{X}, Y_{i-1}, \ldots, Y_1))}{q - \sum_{i=1}^q \xi(Y_i|(Y_{i-1}, \ldots, Y_1))}, \quad \textit{where } T(Y_1|\emptyset) := 0. \tag{3}$$

It can easily be seen that T extends ξ to a dependence measure between multivariate random vectors. First, we observe that T reduces to ξ if $\mathbf{Y}$ is one-dimensional. Further, the numerator in (3) takes values in $[0, q]$ while the

denominator takes values in $[1, q]$. Hence, using that ξ satisfies the information gain inequality, it follows that T only takes values in $[0, 1]$. We observe that T attains the value 0 if and only if $\sum_{i=1}^{q} \xi(Y_i|(\mathbf{X}, Y_{i-1}, \ldots, Y_1)) = \sum_{i=1}^{q} \xi(Y_i|(Y_{i-1}, \ldots, Y_1))$ for all $i \in \{1, \ldots, q\}$. Using again the information gain inequality for ξ, the latter is equivalent to $\xi(Y_i|(\mathbf{X}, Y_{i-1}, \ldots, Y_1)) = \xi(Y_i|(Y_{i-1}, \ldots, Y_1))$ for all i. But this means that Y_i and $\mathbf{X}$ are conditionally independent given $(Y_{i-1}, \ldots, Y_1)$ for all i. Hence, by the chain rule for conditional independence, $\mathbf{X}$ and $\mathbf{Y}$ are independent if and only if $T(\mathbf{Y}|\mathbf{X}) = 0$. Since ξ characterizes perfect directed dependence, it follows similarly that $T(\mathbf{Y}|\mathbf{X}) = 1$ if and only if $\mathbf{Y}$ is perfectly dependent on $\mathbf{X}$. The following result outlines the basic properties of the measure T.

Theorem 1. (Fundamental properties of T; [1]).
The functional T defined by (3) is a dependence measure. Further,

(i) T fulfils the information gain inequality, i.e., $T(\mathbf{Y}|\mathbf{X}) \leq T(\mathbf{Y}|(\mathbf{X}, \mathbf{Z}))$ for all $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, and
(ii) T characterizes conditional independence, i.e., $T(\mathbf{Y}|\mathbf{X}) = T(\mathbf{Y}|(\mathbf{X}, \mathbf{Z}))$ if and only if $\mathbf{Y}$ and $\mathbf{Z}$ are conditionally independent given $\mathbf{X}$.

By the definition in (3), the extension T of Chatterjee's rank correlation is a direct extension based on a dimension reduction where the problem for a q-dimensional random vector is converted into a combination of q one-dimensional problems. The extension T can also be defined for different permutations of the components of $\mathbf{Y}$. To this end, denote by S_q the set of permutations of $\{1 \ldots, q\}$. Then, define $T_\sigma(\mathbf{Y}|\mathbf{X}) := T(\mathbf{Y}_\sigma|\mathbf{X})$ where $\mathbf{Y}_\sigma = (Y_{\sigma_1}, \ldots, Y_{\sigma_q})$ for $\sigma = (\sigma_1, \ldots, \sigma_q) \in S_q$. Averaging over several permutations $S \subseteq S_q$, $S \neq \emptyset$, we define

$$T_S(\mathbf{Y}|\mathbf{X}) := \frac{1}{|S|} \sum_{\sigma \in S} T_\sigma(\mathbf{Y}|\mathbf{X}).$$

It can easily be seen that the properties of T in Theorem 1 also hold for T_S, see [1]. Fast and model-free estimators for T and T_S can be obtained by plugging in the nearest-neighbor estimator ξ_n in (2).

The second recently studied extension of Chatterjee's rank correlation is the kernel partial correlation coefficient defined on topological spaces, see [4]. It is based on the *maximum mean discrepancy* (MMD), which is defined for probability measures Q_1 and Q_2 by the distance $\mathrm{MMD}(Q_1, Q_2) := \|\mu_{Q_1} - \mu_{Q_2}\|_{\mathcal{H}_\mathcal{Y}}$ where $\|\cdot\|_{\mathcal{H}_\mathcal{Y}}$ denotes the norm of a reproducing kernel Hilbert space $\mathcal{H}_\mathcal{Y}$ with uniquely determined kernel $k\colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ on a topological space $\mathcal{Y}$, and μ_{Q_i} is the mean element of Q_i, $i = 1, 2$. An alternative representation is $\mathrm{MMD}^2(Q_1, Q_2) = \mathbb{E}[k(S, S')] + \mathbb{E}[k(T, T')] - 2\mathbb{E}[k(S, T)]$ for $S, S' \sim Q_1$ independent and $T, T' \sim Q_2$ independent. We refer to the literature for details. Denote by $P^{\mathbf{Y}}$ ($P^{\mathbf{Y}|\mathbf{X}}$) the distribution of $\mathbf{Y}$ (conditional on $\mathbf{X}$).

Definition 2. (Kernel partial correlation coefficient; [4]).
The kernel partial correlation coefficient depending on the kernel k is defined by

$$\rho_k(\mathbf{Y}|\mathbf{X}) := \frac{\mathbb{E}[\mathrm{MMD}^2(P^{\mathbf{Y}|\mathbf{X}}, P^{\mathbf{Y}})]}{\mathbb{E}[\mathrm{MMD}^2(\delta_{\mathbf{Y}}, P^{\mathbf{Y}})]}\,, \tag{4}$$

where $\delta_{\mathbf{Y}}$ denotes the Dirac measure in $\mathbf{Y}$.

A kernel is said to be *characteristic* if the kernel mean embedding is injective, i.e., $\mu_{Q_1} = \mu_{Q_2} \Rightarrow Q_1 = Q_2$. This equivalently means that $\mathbb{E}_{S\sim Q_1}[f(S)] = \mathbb{E}_{T\sim Q_2}[f(T)]$ for all $f \in \mathcal{H}_Y$ implies $Q_1 = Q_2$.

Theorem 2. (Fundamental properties of ρ_k; [4]).
Assume that $\mathcal{H}_{\mathbf{Y}}$ is separable, $k(\cdot,\cdot)$ is characteristic, and $\mathbb{E}[k(\mathbf{Y},\mathbf{Y})] < \infty$. Then the kernel partial correlation defined by (4) is a dependence measure. Further,

(i) ρ_k fulfils the information gain inequality, i.e., $\rho_k(\mathbf{Y}|\mathbf{X}) \le \rho_k(\mathbf{Y}|(\mathbf{X},\mathbf{Z}))$ for all $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, and
(ii) ρ_k characterizes conditional independence, i.e., $\rho_k(\mathbf{Y}|\mathbf{X}) = \rho_k(\mathbf{Y}|(\mathbf{X},\mathbf{Z}))$ if and only if $\mathbf{Y}$ and $\mathbf{Z}$ are conditionally independent given $\mathbf{X}$.

Well-known examples of characteristic kernels are the *Gaussian kernel* $k(u,v) = \exp(-\sigma\|u-v\|^2)$, $\sigma > 0$, and the *Laplace kernel* $k(u,v) = \exp(-\sigma\|u-v\|_1)$, $\sigma > 0$, where $\|\cdot\|_1$ denotes the l_1-norm. For the *linear kernel* $k(u,v) = uv$, which is not characteristic, the kernel partial correlation reduces to the Pearson correlation.

3 Multi-output Variable Selection

Both measures T and ρ_k allow a straightforward extension of the feed forward variable selection method established in [2]. The basic idea for the variable selection method is to add step by step that predictor variable to the selected set of predictors which increases the underlying measure the most. This is where the information gain inequality comes in, which ensures that the dependence measure does not decrease when a variable is added to the selected set. The algorithm stops if no predictor variable is left or none of the remaining variables increases the underlying measure given the variables of the selected set. Due to the characterization of conditional independence, the remaining predictor variables are conditionally independent of $\mathbf{Y}$ given the selected variables. Hence, it is sufficient to predict and model $\mathbf{Y}$ based on the selected variables. Note that the estimator satisfies all the appealing properties of the dependence measure only asymptotically.

In this section, we compare the variable selection based on T and ρ_k with each other as well as with a classical Lasso regression in their ability to select relevant factor variables for explaining stock returns.

3.1 Data

As classical factor variables $\mathbf{X}$, we use the Fama French five-factors, i.e., the market factor Mkt-RF, the size factor SMB, the value factor HML, the profitability factor RMW, and the investment factor CMA. The data is taken from the Kenneth French data library.[1] We add placebo factors consisting of arbitrary i.i.d samples from three independent random variables distributed uniformly on $[0, 1]$, standard Gaussian, and exponentially with intensity 1, respectively. Formally, we let

$$\mathbf{X} = (\text{Mkt-RF}, \text{SMB}, \text{HML}, \text{RMW}, \text{CMA}, \text{PLA-Unif}, \text{PLA-Gauss}, \text{PLA-Exp}).$$

Notations and brief descriptions of the factors are summarized in Table 1.

Table 1. The Fama French five-factors and the placebo factors.

Notation	Description
Mkt-RF	Market return minus risk-free rate
SMB	Small minus big
HML	High minus low
RMW	Robust minus weak
CMA	Conservative minus aggressive
PLA-Unif	Placebo factor with uniform distribution on $[0, 1]$
PLA-Gauss	Placebo factor with standard Gaussian distribution
PLA-Exp	Placebo factor with Exp(1)-distribution

As a response variable, we consider a multivariate random vector $\mathbf{Y}$ defined as the daily returns of the big tech companies Apple (AAPL), Amazon (AMZN), Google (GOOGL), Nvidia (NVDA), Meta (META) and Microsoft (MSFT), i.e.

$$\mathbf{Y} = (\text{AAPL}, \text{AMZN}, \text{GOOGL}, \text{NVDA}, \text{META}, \text{MSFT}).$$

The daily stock returns are retrieved as the *TR.TotalReturn1D* variable from Thomson Reuters Eikon database. An overview of the empirical correlations of the big tech companies with the Fama French five-factors and the three placebo factors is given in Table 2.

The joint dataset of factors and response variables ranges from January 2013 to March 2024 and encompasses a total of $n = 2809$ daily observations.

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_5_Factors_2x3_daily_CSV.zip

Table 2. Correlations of the big tech companies with the Fama French five-factors and the three placebo factors.

	Mkt-RF	SMB	HML	RMW	CMA	PLA-Unif	PLA-Gauss	PLA-Exp
Mkt-RF	1.000	0.201	−0.053	−0.198	−0.301	0.012	−0.032	0.002
SMB	0.201	1.000	0.233	−0.273	0.005	0.003	0.002	−0.017
HML	−0.053	0.233	1.000	0.328	0.616	−0.016	−0.004	−0.033
RMW	−0.198	−0.273	0.328	1.000	0.311	0.003	−0.006	−0.044
CMA	−0.301	0.005	0.616	0.311	1.000	−0.040	−0.003	−0.003
MSFT	0.758	−0.067	−0.271	−0.093	−0.405	0.017	−0.035	0.013
AAPL	0.703	−0.012	−0.229	−0.054	−0.292	0.016	−0.011	−0.012
NVDA	0.645	0.075	−0.293	−0.211	−0.425	−0.011	−0.008	0.013
AMZN	0.616	−0.017	−0.344	−0.211	−0.501	0.029	−0.014	−0.001
META	0.574	0.034	−0.264	−0.186	−0.419	0.037	−0.003	−0.001
GOOGL	0.714	0.010	−0.241	−0.139	−0.426	0.004	0.012	−0.020

3.2 Methods

The algorithm for the multi-output variable selection based on T is described in [1, Sect. 6] and referred to as *MFOCI* (*multivariate feature ordering by conditional independence*). For MFOCI, we once use a shuffled set of permutations by letting

$$\bar{S} = \{\sigma \in S_q \mid \sigma = (i+1, \ldots, q, 1, \ldots, i) \text{ for some } i \leq q\},$$

and once the fixed permutation $\bar{\sigma}$ that yields

$$Y_{\bar{\sigma}} = (\text{MSFT}, \text{AMZN}, \text{AAPL}, \text{GOOGL}, \text{META}, \text{NVDA})$$

This permutation $\bar{\sigma}$ is chosen such that the companies are sorted by their correlation with the SMB factor. The shuffled approach with $\bar{S}$ allows to benefit from the averaging effectively while keeping the computational effort at a reasonable level. The latter factor-ordered approach serves as a consistency check that the order of variables does not make a specific factor more prone for selection.

The *KFOCI* (*kernel feature ordering by conditional independence*) method is based on the kernel partial correlation coefficient as introduced in [4]. As noted above, there are different possible choices for the kernel function. In our analysis, we consider the default Gaussian and the Laplace kernel.

Lasso regression is a popular method for (multivariate) variable selection in linear models, and we use it as a benchmark to compare the new model-free methods MFOCI and KFOCI with.

3.3 Implementation

We implement the MFOCI algorithm and make it available as a Python package.[2] The package also includes a demo script that produces similar results

[2] https://pypi.org/project/mfoci/.

to this study based on publicly available data from Yahoo Finance. For the KFOCI variable selection, the implementation is given in the R package `KPC`.[3] The Lasso regression is available in Python's `scikit-learn` package. We leverage the `MultiTaskLassoCV` class to perform the Lasso regression with multiple outputs and a cross-validated regularization parameter.

3.4 Empirical Results

The daily returns of the big tech companies are explained by multiple factors, according to the variable selection results in Tables 3 and 4. In Table 3, all models are successful in selecting the market factor as the most important factor. The CMA factor also plays a prominent role, being selected as the most important factor alongside the market factor in all but one case. This is reasonable in the light of Table 2 showing that the CMA factor also has the highest correlations with the returns of the big tech companies.

Table 3. Factor variable selection for the big tech companies.

Model	Selected factor variables	Coefficients
Shuffled MFOCI	Mkt-RF, CMA, RMW	0.106, 0.143, 0.15
SMB-sorted MFOCI	Mkt-RF, CMA, RMW, HML, PLA-Exp	0.113, 0.145, 0.151, 0.156, 0.159
Gaussian KFOCI	Mkt-RF, HML, SMB, CMA	–
Laplace KFOCI	Mkt-RF, CMA, HML, SMB	–
Lasso	Mkt-RF, CMA, HML, SMB, RMW, PLA-Gauss, PLA-Exp	7.044, 3.874, 2.26, 1.167, 0.793, 0.081, 0.049

Table 4. Factor variable selection for the big tech companies excluding the market factor.

Model	Selected factor variables	Coefficients
Shuffled MFOCI	CMA, RMW	0.044, 0.047
SMB-sorted MFOCI	CMA, RMW	0.032, 0.045
Gaussian KFOCI	CMA, PLA-Gauss, RMW	–
Laplace KFOCI	CMA, RMW, HML, SMB, PLA-Exp	–
Lasso	CMA, RMW, HML, SMB, PLA-Gauss, PLA-Exp	10.306, 0.711, 0.633, 0.45, 0.041, 0.029

[3] https://cran.r-project.org/web/packages/KPC/index.html.

The MFOCI models prefer the RMW factor over the HML factor, despite its smaller correlation with the big tech stock returns. A possible reason is that HML's explanatory power gets partially absorbed by the heavily correlated CMA factor. The SMB factor is of little relevance according to the correlations in Table 2, and it is not selected in the SMB-sorted MFOCI model This underlines that while the order of $\mathbf{Y}$ may affect the selection of factors, the results are not directly linked to an intuitive factor ordering. The placebo factors are generally of little relevance in the selected factors, but neither MFOCI, nor KFOCI, nor Lasso is able to successfully exclude them from the selected factors altogether. One may question if the relevance of other factors increases if the dominant market factor is excluded. Table 4 shows that there is no striking difference in the number of selected factors between the models with and without the market factor, which indicates that the models do not tend to select more irrelevant factors when no factor with a strong signal is present.

In this empirical analysis, the factor variable selection of the MFOCI models is the most consistent across the two tables and the different considered permutations. Note also that whilst the CMA factor coefficient is much larger for the Lasso regression without the market return, such an increase in relevance of a factor does not occur in the MFOCI model results. If we consider the differences

$$T_{\bar{S}}(\text{Mkt-RF}, \text{CMA}|\mathbf{X}) - T_{\bar{S}}(\text{Mkt-RF}|\mathbf{X}) = 0.143 - 0.106 = 0.037,$$
$$T_{\bar{S}}(\text{Mkt-RF}, \text{CMA}, \text{RMW}|\mathbf{X}) - T_{\bar{S}}(\text{Mkt-RF}|\mathbf{X}) = 0.15 - 0.106 = 0.044,$$

the resulting values for MFOCI are only slightly smaller than the shuffled MFOCI coefficients from Table 4, and likewise for the SMB-sorted MFOCI model. This indicates that the market factor does not absorb much of the explanatory power of the other factors in a model-free setting, underlining the uniqueness and usefullness of the other Fama-French factors beyond their correlation with the market return.

4 Conclusion

In this study, we have compared the novel variable selection methods MFOCI and KFOCI as well as the classical Lasso regression in their factor variable selection for a group of stock returns. The models generally agree on the primary factor, but are unable to reliably exclude the placebo factors. In comparison to the Lasso regression, the MFOCI and KFOCI methods select fewer factors, which may help in selecting a more parsimonious model. The results highlight that shuffling the permutations in the MFOCI method is practically relevant for the variable selection and that the choice of the kernel function affects the results of the KFOCI method. Hence, it remains important to choose the exact model setup carefully, which must also be taken into account when interpreting the variables ultimately selected.

Acknowledgements. The first author gratefully acknowledge the support of the Austrian Science Fund (FWF) project P 36155-N ReDim: Quantifying Dependence via Dimension Reduction and the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

References

1. Ansari, J., Fuchs, S.: A simple extension of Azadkia and Chatterjee's rank correlation to a vector of endogenous variables. Preprint arXiv:2212.01621v2 (2022)
2. Azadkia, M., Chatterjee, S.: A simple measure of conditional dependence. Ann. Stat. **49**(6), 3070–3102 (2021)
3. Dette, H., Siburg, K.F., Stoimenov, P.A.: A copula-based non-parametric measure of regression. Scand. J. Statist. **40**(1), 21–41 (2013)
4. Huang, Z., Deb, N., Sen, B.: Kernel partial correlation coefficient—a measure of conditional dependence. J. Mach. Learn. Res. **23**(126), 1–58 (2022)
5. Lin, Z., Han, F.: Limit theorems of Chatterjee's rank correlation. Preprint arXiv:2204.08031 (2022)

Minimizing the Expected Absolute Difference with a Fixed Continuous Copula

Juan Baz[1(✉)], Irene Daz[2], and Susana Montes[1]

[1] Department of Statistics and O.R and M.D., University of Oviedo, Calle Leopoldo Calvo Sotelo 18, Oviedo, Spain
bazjuan@uniovi.es, montes@uniovi.es

[2] Department of Computer Science, University of Oviedo, Calle Pedro Puig Adam, Gijón, Spain
sirene@uniovi.es

Abstract. Given a fixed random variable X and a fixed continuous copula C, the existence of a random variable Y such that (X, Y) has copula C and the quantity $E[|X - Y|]$ is minimum is proved. The proof, based on the Berge Maximum Theorem, also gives the expression of the distribution function of Y and implies a monotonicity property. In addition, the explicit solution for some of the most relevant copulas is stated.

Keywords: Copulas · Distance measure · Probability · Gini's mean difference

1 Introduction

In recent papers, see [4,8], a minimization problem related with the Gini's mean difference [6] has been studied in detail. In particular, given a random variable X and a bivariate copula C, they look for a random variable Y such that the random vector (X, Y) has copula C and the quantity $E[|X - Y|]$ is as small as possible. The quantity $E[|X - Y|]$ can be seen as a distance measure between X and Y, while its minimum as a variability measure of X [5].

As stated in [8], $E[|X - Y|]$ can be expressed as

$$E[|X - Y|] = \int_{-\infty}^{\infty} F(t) + G(t) - 2C(F(t), G(t))dt, \tag{1}$$

where F and G denote, respectively, the distribution function of X and Y.

In the aforementioned papers, the problem is restricted by considering that $G(t)$ should be of the form $h_\theta(F(t))$, being $(h_\theta, \theta \in \Theta)$ a family of distortion functions, i.e. they are increasing and fulfill $h_\theta(0) = 0$ and $h_\theta(1) = 1$ for any

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 18–26, 2024.
https://doi.org/10.1007/978-3-031-65993-5_3

$\theta \in \Theta$. In addition, several assumptions are made both for C and X. Under these conditions, the existence of the minimum is proved.

In this paper, we consider the problem in a more general scheme. No restrictions are considered for the distribution of X and Y and the copula C is only required to be continuous. By minimizing point by point the integrand in Eq. (1), we obtain a possible expression of $G(t)$. Then, using the Berge Maximum Theorem [3] and properties of copulas we state that the proposed $G(t)$ is indeed a distribution function, proving the existence of the solution.

The rest of the paper is organized as follows. In Sect. 2, the basic notions needed for the development of the paper are given. The construction of the solution for the minimizing problem is studied in Sect. 3. Finally, the explicit solution for some relevant copulas is given in Sect. 4 and the conclusions are discussed in Sect. 5.

2 Preliminaries

In the following, we will consider random variables defined over a standard probability space [1]. We recall that, given a random variable X, its distribution function is defined as $F(t) = P(X \leq t)$ and its quantile function is defined as $F^{-1}(p) = \inf\{x \in \mathbb{R} || \ p \leq F(x)\}$ [9]. A function is a distribution function if and only if is increasing, right continuous and fulfills $\lim_{x\to-\infty} F(t) = 0$ and $\lim_{x\to\infty} F(t) = 1$. We recall that a function $f : \mathbb{R} \to \mathbb{R}$ is right continuous if the limit from the right exists for any $c \in \mathbb{R}$ and equals $f(c)$, that is $\lim_{x\to c+} = f(c)$.

Similarly, given a random vector consisting of two random variables X_1 and X_2, (X_1, X_2), its distribution function is defined as $F(t, s) = P(X_1 \leq t, X_2 \leq s)$. The distribution function of a bivariate vector can be decomposed in the (marginal) distribution functions of X_1 and X_2 and the dependence between them. The dependence can be modeled by using copula functions, defined as follows.

Definition 1. *[7] A function $C : [0,1]^2 \to [0,1]$ is said to be a (bivariate) copula if:*

- $C(0, y) = C(x, 0) = 0$ *for any* $x, y \in [0, 1]$,
- $C(1, x) = C(x, 1) = x$ *for any* $x \in [0, 1]$,
- $C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1)$ *for any* $x_1, x_2, y_1, y_2 \in [0, 1]$ *such that* $x_1 \leq x_2$ *and* $y_1 \leq y_2$.

The most common copulas are the countermonotone $C(x, y) = \max(x + y - 1, 0)$, independent $C(x, y) = xy$ and comonotone $C(x, y) = \min(x, y)$ copulas, associated with perfectly negative dependent, independent and perfect positive dependent random variables [7]. In this paper, we will also consider one of the most known families of copulas, the Farlie-Gumbel-Morgenstern (FGM) copulas, defined as follows.

Definition 2. *[7] A copula C is said to be a Farlie-Gumbel-Morgenstern copula if $C(x, y) = xy + \lambda x(1 - x)y(1 - y)$ for any $x, y \in [0, 1]$, where $\lambda \in [-1, 1]$.*

The most known result involving copulas is the Sklar Theorem, that states that any multivariate distribution function can be seen as the composition of a copula and the marginal distribution functions. Here we present the result for dimension 2.

Theorem 1. *[11] Let (X_1, X_2) be a bivariate random vector with distribution function F and marginal distribution functions F_1 and F_2. Then, there exists a copula C such that $F(t, s) = C(F_1(t), F_2(s))$ for any $s, t \in \mathbb{R}$.*

For the development of our main result, we need some notions of continuity of set-valued functions. In particular, we will use the upper hemicontinuity. Given a set B, the set consisting of all the subsets of B will be denoted as $\mathcal{P}(B)$.

Definition 3. *[2] Given two topological spaces A and B, a function $H : A \to \mathcal{P}(B)$ is said to be upper hemicontinuous if for any $a \in A$ and V a open set such that $H(a) \subseteq V$, there exists a neighborhood of a, U, such that $H(a') \subseteq V$ for any $a' \in U$.*

Here, we provide a simplified version of the Berge Maximum Theorem [3] that gives conditions for which, given a two dimensional function, the set of maximizers in one of the variables is a non-empty, compact and upper hemicontinuous set-function with respect to the other variable.

Theorem 2. *Let $f : [0,1] \times [0,1] \to \mathbb{R}$ a continuous function. Then, the set-valued function defined as $H : [0,1] \to \mathcal{P}([0,1])$ such that*

$$H(x) = \left\{ y \in [0,1] \mid f(x,y) = \max_{z \in [0,1]} f(x,z) \right\} = \arg\max_{z \in [0,1]} f(x,z)$$

has non-empty compact values and is upper hemicontinuous.

The last concept that will be of interest is the usual stochastic order (also known as First Order Stochastic Dominance) between two random variables. There exist, see [10], many equivalent definitions for this notion. Here we provide the one that uses distribution functions.

Definition 4. *[10] Let X_1 and X_2 be two random variables with distribution functions F_1 and F_2. It is said that X_1 is smaller than or equal to X_2 with respect to the usual stochastic order (denoted by $X_1 \leq_{st} X_2$) if $F_1(t) \geq F_2(t)$ for any $t \in \mathbb{R}$.*

Another related notion is the almost surely inequality $X \leq_{a.s} Y \iff P(X \leq Y) = 1$. Trivially, $X \leq_{a.s} Y \implies X \leq_{st} Y$.

3 Minimizing the Expected Absolute Difference with a Fixed Continuous Copula

Considering a random variable X and a copula C, denote as $L_{X,C}$ the set of random variables such that, if $Y \in L_{X,C}$, then (X, Y) has the copula C. Our

objective is to try to find $Y \in L_{X,C}$ such that

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E\left[|X-Z|\right]$$

Firstly we will prove the existence of such random variable.

Theorem 3. *Let C be a continuous copula. Then, given a random variable X, there exists a random variable Y such that (X,Y) has copula C and*

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|]$$

Proof. Firstly, denoting the distribution function of X and Y as F and G, one has (1),

$$E\left[|X-Y|\right] = \int_{-\infty}^{\infty} F(t) + G(t) - 2C\left(F(t), G(t)\right) dt.$$

Since the values of $F(t)$ and $G(t)$ are in the interval $[0,1]$, we define the function $h : [0,1] \to [0,1]$ as

$$h(x) = \max\left(\arg\min_{y \in [0,1]} x + y - 2C(x,y)\right). \tag{2}$$

and then determine $G(t) = h(F(t))$. For this strategy to be valid, the only requirement we need is $h(F(t))$ to be a distribution function for each possible distribution function $F(t)$. In particular, we need to prove h to be well defined, increasing, right-continuous and that it fulfills $h(0) = 0$ and $\lim_{x \to 1} h(x) = 1$.

- h is well defined. Since C is continuous, $x+y+C(x,y)$ is also continuous. Then, we can apply Theorem 2 to $-(x+y+C(x,y))$ to obtain that $\arg\min_{y \in [0,1]} x + y - 2C(x,y)$ is compact for any $x \in [0,1]$. Then, the maximum exists and h is well defined.
- h is increasing. If h is not increasing, there exists $x \in [0,1]$ and $\epsilon, \alpha > 0$ fulfilling $x + \epsilon \leq 1$ and $h(x) - \alpha \geq 0$ such that $h(x+\epsilon) = h(x) - \alpha$. Using the properties of copulas, we have $C(x+\epsilon, h(x)) - C(x, h(x)) - C(x+\epsilon, h(x) - \alpha) + C(x, h(x) - \alpha) \geq 0$.

  Therefore, one has

$$\begin{gathered} x + \epsilon + h(x) - 2C(x+\epsilon, h(x)) - (x + \epsilon + h(x) - \alpha - 2C(x+\epsilon, h(x)-\alpha)) = \\ = x + h(x) - (x + h(x) - \alpha) - 2\left(C(x+\epsilon, h(x)) - C(x+\epsilon, h(x)-\alpha)\right) \leq \\ \leq x + h(x) - (x + h(x) - \alpha) - 2\left(C(x, h(x)) - C(x, h(x)-\alpha)\right) = \\ = x + h(x) - 2C(x, h(x)) - (x + h(x) - \alpha - 2C(x, h(x)-\alpha)) \leq 0 \end{gathered}$$

  where the last equality is reached using $h(x) \in \arg\min_{y \in [0,1]} x + y - 2C(x,y)$. Then, $h(x) - \alpha \neq \max\left(\arg\min_{y \in [0,1]} x + \epsilon + y - 2C(x+\epsilon, y)\right)$ and therefore $h(x+\epsilon) \neq h(x) - \alpha$. It is concluded that h is increasing.

– h is right continuous. Apply again Theorem 2 to obtain that the function $H : [0,1] \to \mathcal{P}([0,1])$ defined as $H(x) = \arg\min_{y\in[0,1]} x + y - C(x,y)$ has compact values and is upper hemicontinuous. In addition, there is the relation $h(x) = \max H(x)$.

  Using the same argument that for proving that h is increasing, it is proved that the values of H are connected, thus they are of the form $[a,b] \subseteq [0,1]$.

  Since h is increasing, for proving that h is right continuous it is enough to prove that for any $x \in [0,1)$ and $\epsilon > 0$ there exists $\delta > 0$ such that $h(x+\delta) - h(x) < \epsilon$.

  Consider the notation $H(x) = [a,b]$ and the open set $(a-\lambda, b+\lambda)$ with $\lambda < \min(a, 1-b, \epsilon)$. Then, using the upper hemicontinuity, there exists a neighborhood of x, U, such that for any $x' \in U$ one has $H(x') \in (a-\lambda, b+\lambda)$. Them, for any $x' \in U$, $h(x') \leq b + \lambda$.

  Since U is a neighborhood of x, there exists $\delta > 0$ such that for any $x' \in [0,1]$ such that $|x - x'| < \delta_0$, $x' \in U$. Therefore, considering $\delta = \frac{\delta_0}{2}$, one has $h(x+\delta) - h(x) \leq b + \lambda - b = \lambda < \epsilon$. It is concluded that h is right continuous.
– $h(0) = 0$. If $x = 0$, then $0 + h(0) - 2C(0, h(0)) = h(0)$ takes the minimum value if and only if $h(0) = 0$.
– $\lim_{x\to 1} h(x) = 1$. Since $h(x)$ is increasing and bounded, the limit $\lim_{x\to 1} h(x)$ exists. Suppose that $\lim_{x\to 1} h(x) = x_0 < 1$. Using again that h is increasing, one has that $h(x) \leq x_0$ for any $x \in [0,1]$. Now, let $x = 1 - \delta$ with $\delta < \frac{1-x_0}{2}$. Then,

$$1 - \delta + h(1-\delta) - 2C(1-\delta, h(1-\delta)) \geq 1 - \delta + h(1-\delta) - 2C(1, h(1-\delta)) = \\ = 1 - \delta - h(1-\delta) > 1 - \delta - x_0 > \tfrac{1-x_0}{2}$$

  In addition, if we replace $h(1-\delta)$ by 1:

$$1 - \delta + 1 - 2C(1-\delta, 1) \geq 1 - \delta + 1 - 2(1-\delta) = \delta < \frac{1-x_0}{2}$$

  This is a contradiction to the fact that $h(1-\delta) < x_0$, since if we replace its value for 1 the value of the objective function is smaller. It is concluded that $\lim_{x\to 1} h(x) = 1$. □

Notice that the proof of Theorem 3 is constructive, we give the solution of the problem. In particular, if X has distribution function $F(t)$, the distribution function of Y is determined as $h(G(t))$, with h defined as (2), i.e. the maximum of the (compact) set of minimizers of the objective function point by point. In addition, the function h only depends on the copula and is increasing. Using these properties, we can prove a monotonicity result.

Proposition 1. *Considering the same conditions as in Theorem 3. If X_1 and X_2 are two random variables such that $X_1 \leq_{st} X_2$, there exist Y_1 and Y_2 two random variables such that:*

$$E[|X_1 - Y_1|] = \min_{Z \in L_{X_1,C}} E[|X_1 - Z|]$$
$$E[|X_2 - Y_2|] = \min_{Z \in L_{X_2,C}} E[|X_2 - Z|]$$

and $Y_1 \leq_{st} Y_2$.

Proof. Denote $F_1(t)$ and $F_2(t)$ the distribution functions of X_1 and X_2. Since $X_1 \leq_{st} X_2$, $F_1(t) \geq F_2(t)$ for any $t \in \mathbb{R}$. As proved in Theorem 3, the distribution functions of Y_1 and Y_2 can be chosen to be $h(F_1(t))$ and $h(F_2(t))$ with h being increasing. Then, one has $h(F_1(t)) \geq h(F_2(t))$ for any $t \in \mathbb{R}$ and it is concluded that $Y_1 \leq_{st} Y_2$. □

4 Solutions for Some Particular Copulas

We devote this section to give the solutions for some cases of interest. Let us start considering the comonotone copula. Since in this case the random variables X and Y have a perfect positive dependence, trivially the minimum of $E[|X - Y|]$ is reached when $X =_{a.s.} Y$. However, we can prove this statement also using our technique.

Proposition 2. *Let C be the minimum copula. Then, given a random variable X, for any random variable Y such that $X =_{a.s} Y$ one has that*

$$E[|X - Y|] = \min_{Z \in L_{X,C_{\min}}} E[|X - Z|].$$

Proof. The function we have to minimize is $x+y-2\min(x,y)$. If we derive it with respect to y, we obtain -1 if $y < x$ and 1 if $y > x$. Therefore, since the function is continuous, the minimum is reached when $y = x$. Then, we define $h(x) = x$, thus Y has the same distribution as X and it is concluded that $X =_{a.s} Y$. □

However, for the other two main copulas the solution is not that straightforward. In particular, the best choice is to consider a degenerate random variable that takes the median of X as its value.

Proposition 3. *Let C be the independence or the countermonotone copula. Then, given a random variable X with distribution function $F(t)$, for any random variable Y such that $Y =_{a.s.} F^{-1}(0.5)$, one has that*

$$E[|X - Y|] = \min_{Z \in L_{X,C}} E[|X - Z|]$$

Proof. For the independence copula, the function we have to minimize is $x+y-2xy$. If we derive it with respect to y, we obtain $1-2x$, which is positive when $x < 0.5$, negative when $x > 0.5$ and 0 if $x = 0.5$. Then, h can be considered to be $h(x) = 0$ if $x < 0.5$ and $h(x) = 1$ otherwise. Then, $h(F(t))$ corresponds to a degenerate random variable with value $F^{-1}(0.5)$.

For the countermonotone copula, the function we have to minimize is $x+y-2\max(x+y-1,0)$. If we derive it with respect to y, we obtain 1 if $x+y<1$ and -1 otherwise. Then, the minimum is reached in $y=0$ or in $y=1$. In the first case, one has that the function is $x+0-2\max(x+0-1,0)=x$ and in the second one $x+1-2\max(x+1-1,0)=1-x$, thus the minimum is reached in $y=0$ when $x<0.5$, in $y=1$ when $x<0.5$ and in both ends when $x=0.5$. For the rest of the steps, proceed as the previous case. □

Since the median is the solution for the independence and the countermonotone copula, a natural question is to consider if this also holds for any copula between them. In particular, since the countermonotone copula is the smallest one, it is equivalent to prove the following result.

Proposition 4. *Let C be a copula such that $C(x,y) \leq xy$ for all $x,y \in [0,1]$. Then, given a random variable X with distribution function $F(t)$, for any random variable Y such that $Y =_{a.s.} F^{-1}(0.5)$ on has that*

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|]$$

Proof. Denote as C_1 and C_2 the independent and countermonotone copulas. Since C_2 is the smallest copula, one has that $C_2(x,y) \leq C(x,y) \leq C_1(x,y)$ for any $x,y \in [0,1]$. Then, we can express $C(x,y) = \lambda(x,y)C_1(x,y) + (1-\lambda(x,y))C_2(x,y)$ with $\lambda(x,y) \in [0,1]$ for any $x,y \in [0,1]$. Then, the function to minimize is of the form

$$\begin{aligned} x+y-2C(x,y) &= x+y-2\left(\lambda(x,y)C_1(x,y)+(1-\lambda(x,y))C_2(x,y)\right) = \\ &= \lambda(x,y)\left(x+y-C_1(x,y)\right)+(1-\lambda(x,y))\left(x+y-C_2(x,y)\right) \end{aligned}$$

Using Proposition 3. the minimum of $(x+y-C_1(x,y))$ and $(x+y-C_2(x,y))$ for any $x \in [0,1]$ is reached in the same point, $y=0$ if $x<0.5$ and $y=1$ otherwise. Then, regardless of the value of $\lambda(x,y)$, the minimum of $x+y-2C(x,y)$ is also reached in $y=0$ when $x<0.5$ and $y=1$ otherwise. Then, define h as in the proof of Proposition 3 to conclude the proof. □

Notice that the latter result holds for any copula (not necessarily continuous) that is smaller than the independence copula. We end this section by providing the solution for the family of FGM copulas.

Proposition 5. *Let C be a FGM copula with parameter $\lambda \in [-1,1]$. Then, given a random variable X with distribution function $F(t)$, for any random variable Y such that,*

- $Y =_{a.s.} F^{-1}(0.5)$ *if* $\lambda \leq 0$,
- *Its distribution function is* $h(F(t))$ *with* $h(x)$ *being the median value between* 0, $\frac{1}{2}\frac{2\lambda x(1-x)-1+2x}{2\lambda x(1-x)}$ *and* 1 *(when* $x \notin \{0,1\}$*) if* $\lambda > 0$,

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|]$$

Proof. For $\lambda \leq 0$, the solution is given by Proposition 4.

Suppose that $\lambda > 0$. The function we have to minimize is $x+y-2xy-2\lambda x(1-x)y(1-y)$. If we derive it with respect to y, we obtain $1-2x-2\lambda x(1-x)(1-2y)$. Moreover, if we derive again the function, we obtain $4\lambda x(1-x)$, which is positive for any $x \in [0,1]$.

The solution (for y) of $1-2x-2\lambda x(1-x)(1-2y)=0$ is $y = \frac{1}{2}\frac{2\lambda x(1-x)-1+2x}{2\lambda x(1-x)}$. If $y \in [0,1]$, then is the minimum. If $y < 0$, since the second derivative is positive, the derivative is positive in $[0,1]$ and therefore the minimum over $[0,1]$ is located in 0. Similarly, if $y > 1$, the minimum over $[0,1]$ is located in 1. □

5 Conclusions

In this paper, the problem of minimizing the absolute difference between two random variables with a fixed copula is studied. In particular, it is proved that, when the copula is continuous, the minimum exists and the distribution of a minimizer is provided. In addition, a monotonicity property and the solution for the comonotone, independent, countermonotone and FGM copulas is stated.

The main open question is to study the problem in the more general setting of non-continuous copulas. We do not believe that the techniques used in this paper can be extended for that case, since the Berge Maximum Theorem needs the function to be continuous. Another interesting question is if these results can be extended for random vectors.

Acknowledgements. The authors have been supported by Programa Severo Ochoa of Principality of Asturias (BP21042) and the Ministry of Science and Innovation (PID2022-139886NB-l00)

References

1. Athreya, K.B., Lahiri, S.N.: Measure Theory and Probability Theory. Springer Texts in Statistics, vol. 19. Springer, New York, NY (2006)
2. Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Modern Birkhäuser Classics. Birkhäuser Boston, MA (2009)
3. Berge, C.: Topological Spaces: Including a Treatment of Multi-Valued Functions. Vector Spaces and Convexity, Oliver and Boyd (1963)
4. Capaldo, M., Di Crescenzo, A., Pellerey, F.: Generalized Gini's mean difference through distortions and copulas, and related minimizing problems. Statist. Probab. Lett. **206**, 109981 (2024)

5. Capaldo, M., Navarro, J.: New multivariate Gini's indices. Preprint arXiv:2401.01980 (2024)
6. David, H.A.: Miscellanea: Gini's mean difference rediscovered. Biometrika **55**(3), 573–575 (1968)
7. Nelsen, R.B.: An Introduction to Copulas. Springer (2006)
8. Ortega-Jiménez, P., Pellerey, F., Sordo, M.A., Suárez-Llorens, A.: A minimizing problem of distances between random variables with proportional reversed hazard rate functions.In: García-Escudero, L.A., et al. (eds.) Building Bridges between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing, vol. 1433, pp. 311–318. Springer, Cham (2023)
9. Rohatgi, V.K., Saleh, A.M.E.: An Introduction to Probability and Statistics. Wiley (2015)
10. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer (2007)
11. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)

An EM Stopping Rule for Avoiding Degeneracy in Gaussian-Based Clustering with Missing Data

Christophe Biernacki[1(✉)] and Vincent Vandewalle[2]

[1] Inria, Université de Lille, CNRS, UMR 8524-Lab. P. Painlevé, 59000 Lille, France
christophe.biernacki@inria.fr
[2] Université Côte d'Azur, Inria, CNRS, UMR 7351-Lab. J.A. Dieudonné, 06108 Nice, France
vincent.vandewalle@inria.fr

Abstract. Missing data frequency increases with the growing size of multivariate modern datasets. In Gaussian model-based clustering, the EM algorithm easily takes into account such data but the degeneracy problem is dramatically aggravated during the EM runs: parameter degeneracy is quite slow and also more frequent than with complete data. Consequently, parameter degenerated solutions may be confused with valuable parameter solutions and, in addition, computing time may be wasted through wrong runs. In this work, a simple and low informational condition on the latent partition allows to propose a very simple partition-based stopping rule of EM which shows good behavior on numerical experiments.

Keywords: Clustering · Degeneracy · EM algorithm · Gaussian mixtures · Latent partition · Missing data

1 Introduction

Gaussian mixtures models easily fit many practical settings, like the clustering paradigm. Various estimation approaches are available, but the maximum likelihood (ML) approach is usually much preferred. Nevertheless, it is well-known that the likelihood function of heteroscedastic normal mixture models is not bounded from above [1,5]. As a consequence, firstly some theoretical questions about the ML properties are raised and, secondly, optimization algorithms like EM [2] may converge, as observed by any practitioner, towards such degenerate solutions. When it encounters a degenerated solution the likelihood goes to infinity, which is a symptom that the parameters are close of the border. These degenerated solutions on the border of the parameters space are obviously out of interest.

The solutions to avoid the problem of degeneracy are either to modify the estimator of the parameters or to detect it through the dynamic of the EM algorithm. Avoiding degeneracy is usually handled by some *parameter-based constraints*, typically by constraining the singular values of the covariance matrices;

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 27–35, 2024.
https://doi.org/10.1007/978-3-031-65993-5_4

see for instance [3,8]. Alternatively, two very different approaches have been proposed. A very recent one is proposed by [7] through the so-called (bounded) marginal likelihood consisting to estimating first the variances (and the mixture proportions) while integrating over all positioning information, and then to estimate the centers conditionally to the previous estimates. Despite its conceptual interest, this approach produces some estimates loosing some Fisherian information (is hardly implemented in the multivariate Gaussian case and for more than two components). A second alternative approach has been proposed three decades ago by [6] and could be qualified as a *partition-based constraint* approach. The author imposes a constraint on the latent partition underlying the data, that leads to maximize a (bounded) conditional likelihood and gives consistent estimates. The proposed assumption is weak and natural since it only requires that at least $d+1$ data units arise from each d-variate mixture Gaussian component.

However, very few results are available for the degeneracy of Gaussian mixtures with some missing data while, with the increasing of the number of available variables, the risk that data contain missing values also increases. In this article we study the degeneracy properties in the missing data framework. Some unexpected results will be revealed, underlying the necessity to avoid degeneracy with most priority than in the complete data case itself. It will lead to design a interesting solution consisting of an absolutely standard EM algorithm combined with an easy-to-compute partition-based stopping rule.

The outline of this paper is the following. In Sect. 2, we characterize the EM algorithm dynamics in the missing data case within the Gaussian framework. In Sect. 3, we present the partition-based constraint solution for stopping the EM algorithm when engaged along a possible degeneracy run. Section 4 illustrates through empirical studies the efficiency of our proposal. The last section concludes this work.

2 Degeneracy of the EM Algorithm with Missing Data

2.1 The EM Algorithm for Missing Data

Missing Data Let consider a sample $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ coming from a mixture of K Gaussian components in $\mathcal{R}^d$. Let denote by $O_i \subseteq \{1, \ldots, d\}$ the set of observed variables for the individual i ($i \in \{1, \ldots, n\}$) and by M_i the complementary set for the missing variables. Let z_{ik} denote the class of i: $z_{ik} = 1$ if $\mathbf{x}_i$ comes from class k ($k \in \{1, \ldots, K\}$) and 0 otherwise. The whole partition is denoted $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n) \in \mathcal{Z}$, where $\mathbf{z}_i = (z_{i1}, \ldots, z_{iK})$. Let $\mathbf{x}_i^o$ denote the observed variables for unit i, and $\mathbf{x}^o$ the observed data set.

The Mixture Model From the model point of view, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ are the proportions of each component, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and covariance matrix of the class k. Let $\boldsymbol{\theta} \in \Theta$ be the global parameter of the mixture, $\boldsymbol{\pi}$ included. Finally $\phi(\cdot; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian density of expectation $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. μ_{ik}^o is the sub-vector of $\boldsymbol{\mu}_k$ associated to the index O_i (*idem* for M_i). $\boldsymbol{\Sigma}_{ik}^{om}$ is the sub-matrix of $\boldsymbol{\Sigma}_k$ associated to the rows O_i and

columns M_i (and the same for any combination of indexes o and m, like $\mathbf{\Sigma}_{ik}^{oo}$, *etc.*).

The Related EM Algorithm Starting from an initial parameter, the EM algorithm allowing to take into account missing data is described below, where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^+$ denote respectively parameters at two successive iterations (a similar convention is adopted for the missing data $\mathbf{x}^m$ and $\mathbf{z}$). Note also that this algorithm is stopped typically either after a predefined number of iterations, or when a predefined threshold in the increase of the log-likelihood is reached.

E step Missing data estimation through the computation of $t_{ik}^+ \propto \pi_k \phi(\mathbf{x}_i^o; \boldsymbol{\lambda}_k)$ and $\mathbf{x}_{ik}^{m^+} = \boldsymbol{\mu}_{ik}^m + \mathbf{\Sigma}_{ik}^{mo} \left(\mathbf{\Sigma}_{ik}^{oo}\right)^{-1} (\mathbf{x}_i^o - \boldsymbol{\mu}_{ik}^o)$. Here $\mathbf{x}_{ik}^{m^+}$ can be interpreted as an imputation of missing data given the class and the observed variables.

M step Parameter estimation through the computation of $\pi_k^+ = \frac{1}{n_k^+} \sum_{i=1}^n t_{ik}^+$, $\boldsymbol{\mu}_k^+ = \frac{1}{n_k^+} \sum_{i=1}^n t_{ik}^+ \mathbf{x}_{ik}^+$ and $\mathbf{\Sigma}_k^+ = \frac{1}{n_k^+} \sum_{i=1}^n t_{ik}^+ \Big[(\mathbf{x}_{ik}^+ - \boldsymbol{\mu}_k^+)(\mathbf{x}_{ik}^+ - \boldsymbol{\mu}_k^+)' + \mathbf{\Sigma}_{ik}^+ \Big]$ where $n_k^+ = \sum_{i=1}^n t_{ik}^+$, $\mathbf{x}_{ik}^+ = \begin{pmatrix} \mathbf{x}_i^o \\ \mathbf{x}_{ik}^{m^+} \end{pmatrix}$ and $\mathbf{\Sigma}_{ik}^+ = \begin{pmatrix} \mathbf{0}_i^o & \mathbf{0}_i^{om} \\ \mathbf{0}_i^{mo} & \mathbf{\Sigma}_{ik}^{m+} \end{pmatrix}$ with $\mathbf{0}$ being the null matrix $d \times d$ and $\mathbf{\Sigma}_{ik}^{m^+} = \mathbf{\Sigma}_{ik}^{mm} - \mathbf{\Sigma}_{ik}^{mo} \left(\mathbf{\Sigma}_{ik}^{oo}\right)^{-1} \mathbf{\Sigma}_{ik}^{om}$. $\mathbf{\Sigma}_{ik}^{m^+}$ can be interpreted as a variance fitting in order to take into account the under estimation of the variance caused by the missing data imputation.

Unbounded Likelihood Let $\ell(\boldsymbol{\theta}; \mathbf{x}^o) = \sum_{i=1}^n \log \left(\sum_{k=1}^K \pi_k \phi(\mathbf{x}_i^o; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k) \right)$ the log-likelihood. Classically, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is defined by $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \mathbf{x}^o)$. However, if $\mathbf{\Sigma}_k$ is free (heteroscedastic case) then $\ell(\boldsymbol{\theta}; \mathbf{x}^o)$ is unbounded for instance by taking a center such that $\boldsymbol{\mu}_k = \boldsymbol{x}_i$ and the corresponding *generalized variance* such that $|\mathbf{\Sigma}_k| \to 0$ [5].

2.2 Unexpected EM Algorithm Behaviour in Case of Missing Data

Illustration of a Specific EM Dynamics on Real Data EM dynamics is well-identified in the *complete individual data* case [4]. Typically, when an EM iteration is close enough to a degenerate situation for a given component, then the corresponding generalized variance evolves exponentially fast towards zero along the subsequent iterations of the algorithm. However, a very different behavior has been identified in the case of missing data as we illustrate now on the real data set breast cancer tissue of the UCI database repository.[1] It is composed with $n = 106$ statistical units and $d = 9$ features. We have artificially hidden 10% of the data completely at random. Then we have adjusted a mixture model with $K = 4$ clusters. Then, the evolution of the log-likelihood for a degenerated solution is given in Fig. 1 (left panel). We can see that the growth of the log-likelihood seems to be linear. If we look now at the evolution of the logarithm of

[1] https://archive.ics.uci.edu/ml/datasets/Breast+Cancer.

the generalized variance of the degenerating component Fig. 1 (right panel) we also see that it seems to decrease linearly as the number of iteration increases. Such a log-likelihood and a log-generalized variance behavior indicate clearly that the EM dynamics in the missing individual data case is radically different from this one observed in the complete individual data case.

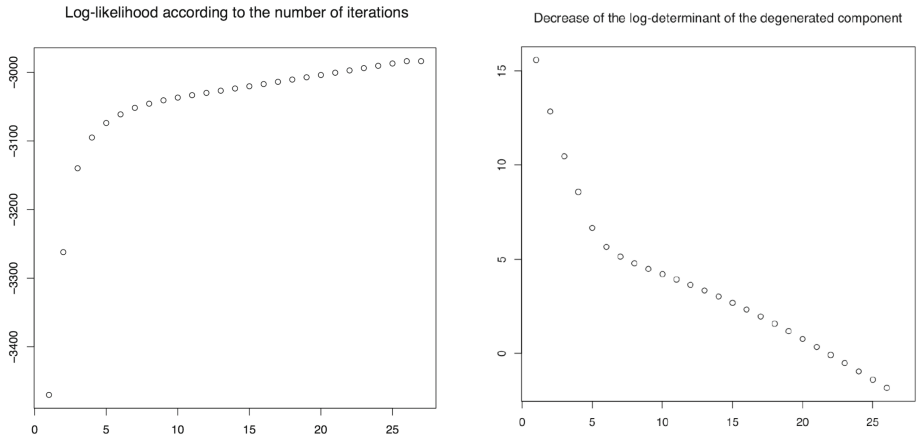


Fig. 1. Illustration of the EM degeneracy dynamics with missing data, according to the number of iterations: (left) Evolution of the likelihood; (right) Evolution of the logarithm of the generalized variance for the degenerating component.

Explaining the Observed Degeneracy Speed Through a Toy Example In order to guess the dynamic of the EM algorithm, let start with a Gaussian univariate framework ($d = 1$) without mixture ($K = 1$) and with only one datum that we will denote simply by x. In this framework the estimator of the mean μ is equal to x and the estimator of the variance Σ is equal to 0 which leads to an infinite likelihood. Let imagine that in addition to this unique observed datum, $n-1$ data have not been observed (they are missing). Then, it is possible to perform an EM algorithm (useless here) which will converge towards the expected degenerated solution. However, in this oversimplified instance, it is possible to explicitly express the algorithm pathway, since updating formulas at the M step become $\mu^+ = n^{-1}((n-1)\mu + x)$ and $\Sigma^+ = n^{-1}((n-1)\Sigma + (x-\mu^+)^2)$. After simple algebra, we can then express at iteration $q \geq 0$ both the convergence speed of the log-likelihood ℓ and of the variance Σ:

$$\ell(\boldsymbol{\theta}^{(q)}; x) \sim -0.5q \ln \frac{n-1}{n} \quad \text{and} \quad \Sigma^{(q)} \sim \Sigma^{(0)} \ln \left(\frac{n-1}{n} \right)^q,$$

where $\boldsymbol{\theta}^{(q)}$ denotes the parameter $\boldsymbol{\theta}$ at iteration q. We thus retrieve in this simple example, the previous experimental behavior of both the log-likelihood and of

the variance along a degeneracy run. Let also remark that the degeneracy speed decreases as the rate of missing data increases. It can thus be expected that the EM algorithm can be trapped by degenerated solutions with a still slower dynamics, as we illustrate now.

Additional Behaviour of EM: Influence of the Missing Data Rate Let take again the instance on the breast cancer tissue of the UCI database repository, and now we vary the rate of missing data. From results presented Table 1 we can see that as the rate of missing data increases, the rate of degeneracy increases, and the number of iterations before degeneracy decreases.

Table 1. Frequency and speed of degeneracy (deg.) according to the rate of missing data on the breast cancer data set.

% missing data	0	5	10	15	20	25	30
% deg	16	4	12	11	46	51	100
Average nb of iterations before deg	2	13	13	82	304	138	215

3 Minimal Partition Information for Avoiding Degeneracy

Usual solutions to avoid degeneracy have been already briefly presented in Sect. 1. Here we focus on the solution consisting of introducing a constraint on the latent partition within the estimation process, arguing that it represents the method with the weakest information required within a model-based clustering context. We present this method below, which has also the advantage to be easily applicable to the missing data case, but by introducing as a novelty a specific adaptation which avoids its associated combinatorics limitations.

3.1 The Partition Information and the Associated EM* Algorithm

The core idea is here to transpose some natural constraints from the supervised setting to the current unsupervised setting, *i.e.* enough data should be available in each class to perform the parameter estimation (an obvious requirement to obtain non-singular covariance matrices). It is equivalent to constrain the partition space $\mathcal{Z}$. This idea was first introduced by [6] and we detail it below.

This strategy being originally introduced for the complete data case, let $n_k = \sum_{i=1}^{n} z_{ik}$ denoting the number of individuals in component k and let define $\mathcal{Z}^* = \{\mathbf{z} : \forall k, n_k \geq d+1\}$ the set of partitions with at least $d+1$ elements by class. In that case, [6] proposed to maximize the *conditional* log-likelihood $\ell(\boldsymbol{\theta}; \mathbf{x}|\mathcal{Z}^*)$, which can be optimized with a specific EM algorithm and which leads

also to consistent mixture estimates. At this step, the key point is to notice that $L(\boldsymbol{\theta}; \mathbf{x}|\mathcal{Z}^*) < \infty$ with probability one. In the following, we however pursue with the *joint* likelihood $L(\boldsymbol{\theta}; \mathbf{x}, \mathcal{Z}^*) = L(\boldsymbol{\theta}; \mathbf{x}|\mathcal{Z}^*)\mathrm{p}(\mathcal{Z}^*)$ since it will be more tractable from the computational point of view in the M step of EM (see further) than its conditional version and both joint and conditional likelihoods are extremely similar ($L(\boldsymbol{\theta}; \mathbf{x}, \mathcal{Z}^*) \simeq L(\boldsymbol{\theta}; \mathbf{x}|\mathcal{Z}^*)$) because $\mathrm{p}(\mathcal{Z}^*) \simeq 1$ even for moderate values of n.

In the missing data case, we have obviously to slightly adapt the definition of the so-called partition information $\mathcal{Z}^*$. Now $\mathcal{Z}^*$ corresponds to a minimum of $d+1$ *complete* individual data in each component, thus namely we define now $\mathcal{Z}^* = \{\mathbf{z} : \forall k, n_k^o \geq d+1\}$, with $n_k^o = \sum_{i:O_i=\{1,\ldots,d\}} z_{ik}$ designating the number of complete individuals in component k. This adaptation is motivated by the fact that the observed likelihood including the partition information $\mathcal{Z}^*$, namely $L(\boldsymbol{\theta}; \mathbf{x}^o, \mathcal{Z}^*)$, is still bounded above with probability one (thus as in the non-missing data case where such a property was straightforward). We make this property clear through the following proposition.[2]

Proposition 1. *For* $\mathcal{Z}^* = \{\mathbf{z} : \forall k, n_k^o \geq d+1\}$ *then* $p(L(\boldsymbol{\theta}; \mathbf{x}^o, \mathcal{Z}^*) < \infty) = 1$.

We adapt now EM for optimizing this particular likelihood, leading to a specific EM algorithm, called here EM*. This latter is expressed as follows:

E step $t_{ik}^{*+} \propto \tau_{ik} t_{ik}^+$ with $\tau_{ik} = \mathrm{p}(\mathcal{Z}^*|\mathbf{x}^o, z_{ik} = 1; \boldsymbol{\theta})$ and $t_{ik}^+ = \mathrm{p}(z_{ik} = 1|\mathbf{x}^o; \boldsymbol{\theta})$.

M step Standard formulas of EM (see Sect. 2.1) where t_{ik}^+ is replaced by t_{ik}^{*+}.

Thus, comparing to the E step of the classical EM, we can observe that a kind of regularization of the conditional probabilities z_{ik}^+ is performed. However, the E step of EM* implies combinatorics within the computation of τ_{ik} as soon as $K > 2$ which greatly limits its practical use (both for the complete and the missing data cases). The following proposed strategy overcomes this problem.

3.2 A Simpler Solution: An Early EM Stopping Rule Relying on $\mathcal{Z}^*$

The term τ_{ik} is the pivotal quantity involved in our strategy. We start by observing two key points in EM*. Firstly, the combinatorial problem only comes from the computation of the term τ_{ik}. Secondly, its value is essentially such that $\tau_{ik} \simeq 1$ (for all i and k values) along most of the iterations of any EM* run, except in a limited number of iterations where the run approaches a degenerate solution. Consequently, most of EM* iterations are greatly equivalent to classical EM iterations (remind that EM and EM* are identical when $\tau_{ik} = 1$), meaning that computing the τ_{ik} values at each iteration of EM* is mostly usefulness. On the contrary, when approaching closely a degenerate solution, we must observe $\tau_{ik} \simeq 0$ for at least one value of the couple (i, k), meaning the corresponding

[2] The proof is not given in this too short paper version.

iteration of EM and EM* diverge, except if simultaneously the value of t_{ik}^{+} is such that $t_{ik}^{+} \simeq 0$.

From this core remark, we formulate a first strategy consisting to replace the calculus of each τ_{ik}, a generally unfeasible task, by a sampling of a partition value $\mathbf{z}_{|ik}$ drawn from $\mathrm{p}(\mathbf{z}|\mathbf{x}^o, z_{ik} = 1; \boldsymbol{\theta})$, which is on the contrary an excessively simple task. Then we simply check if the generated partition value $\mathbf{z}_{|ik}$ belongs or not to $\mathcal{Z}^*$. Indeed, if $\tau_{ik} \simeq 0$, then it is expected to draw $\mathbf{z}_{|ik} \notin \mathcal{Z}^*$, whereas if $\tau_{ik} \simeq 1$, then it is expected to draw $\mathbf{z}_{|ik} \in \mathcal{Z}^*$. Then, in practice, a run of the EM algorithm is stopped as soon as $\mathbf{z}_{|ik} \notin \mathcal{Z}^*$ for at least one couple (i, k), where $\mathbf{z}_{|ik}$ is drawn from $\mathrm{p}(\mathbf{z}|\mathbf{x}^o, z_{ik} = 1; \boldsymbol{\theta})$. Thus, this strategy acts as a partition stopping rule and produces a complete restart of EM from a new run.

However, this first strategy is expected to be too stringent since it could stop some runs where $z_{ik} \simeq 0$ for some (i, k) values, situation where $\tau_{ik} t_{ik}^{+} \simeq t_{ik}^{+}$ even if $\tau_{ik} \simeq 0$. In other words, we could stop some EM runs which are finally equivalent to EM* runs, whereas we target to stop EM runs only in situations where there are different from the EM* ones. It thus leads to the second (and finally retained) strategy that we describe now and which should be preferred to avoid stopping valid EM runs. The principle is to consider the product $\tau_{ik} t_{ik}^{+}$ as a whole, instead of τ_{ik} individually. In that way, the strategy consists to firstly draw a value z_{ik} from t_{ik}^{+}, for each couple (i, k) (thus a full partition $\mathbf{z}$ from $\mathrm{p}(\mathbf{z}|\mathbf{x}^o; \boldsymbol{\theta})$). Then, only for each couple (i, k) s.t. $z_{ik} = 1$, perform the 1st strategy previously described.

Notice that, taken in its globality, this 2nd strategy is equivalent to the very simple procedure consisting finally to draw a $\mathbf{z}$ value from $\mathrm{p}(\mathbf{z}|\mathbf{x}^o; \boldsymbol{\theta})$ and then to check if $\mathbf{z} \in \mathcal{Z}^*$ (in that way EM is not stopped at this current iteration) or $\mathbf{z} \notin \mathcal{Z}^*$ (in that way EM is immediately stopped at this iteration and re-run from another starting value).

4 Numerical Experiments

The proposed strategy is below compared to EM* (used as the reference algorithm and tractable for $K = 2$) with respect to the adjusted rand index (ARI) values between the obtained partition and the simulated partition.

Let consider 100 data sets generated from a 9-variate ($d = 9$) Gaussian mixture of two components. Classes have the same proportions ($\pi_1 = \pi_2 = 0.5$), the covariance matrices are the identity and the class centers are $\mu_1 = (0, 0, ..., 0)'$ and $\mu_2 = (6/\sqrt{d}, 6/\sqrt{d}, ..., 6/\sqrt{d})'$. The number of data n is equal to $n = 150$, and the probability for a data-cell to be missing equals to 0.2. This setting allows having the same Mahalanobis distance between cluster whatever the value. Moreover, the high separation between clusters implies that poor ARI solutions are mainly caused by a poor optimum of the likelihood.

Then, for each given dataset each algorithm is initialized once, with the same initialization for all the algorithms, and a total of 100 iterations. We then compare the ARI resulting from EM and EM*. For each data set, results are split considering three possible cases: first the stopping rule is not activated

(thus in this case degeneracy cannot occur), second the stopping rule has been activated, but no degeneracy has been observed, third the stopping rule has been activated and a degeneracy due to numerical crash has been reported. Results are presented in Table 2. We observe that when the stopping rule is not activated, both EM and EM* give very close ARI values. In addition, when the stopping rule is activated, it appears to be justified either due to a degenerate run, or due to a poor partitioning output value run. Notice these latter runs should possibly correspond to a degenerate run, but it is difficult to access to this information since we know that degeneracy could be quite very slow in the missing data case.

Table 2. Average ARI for EM and EM* (the frequency of each case is also given).

Stopping rule (frequency)	EM	EM*
Not activated (94)	0.909	0.915
Activated without EM degeneracy (5)	0.074	0.340
Activated with EM degeneracy (1)	–	0.604

5 Concluding Remarks

We have identified unexpected, and unpleasant, behaviour of the EM algorithm in the case of missing data, potentially leading to poor clustering results and a waste of computing time. We have proposed an easy-to-implement early stopping rule directly on a *classical* EM algorithm, and relying on a minimal partitioning information. The first promising numerical experiments we obtained have now to be confirmed by future numerical experiments.

References

1. Day, N.E.: Estimating the components of a mixture of normal distributions. Biometrika **56**, 463–474 (1969)
2. Dempster, A.P., Laird, N M., Rubin, D.B.: Maximum likelihood from incomplete data (with discussion). J. R. Stat. Soc. Ser. B Stat. Method. **39**, 1–38 (1977)
3. Ingrassia, S., Rocci, R.: Constrained monotone EM algorithms for finite mixture of multivariate Gaussians. Comput. Stat. Data Anal. **51**(11), 5339–5351 (2007)
4. Ingrassia, S., Rocci, R.: Degeneracy of the EM algorithm for the MLE of multivariate Gaussian mixtures and dynamic constraints. Comput. Stat. Data Anal. **55**(4), 1715–1725 (2011)
5. Kiefer, J., Wolfowitz, J.: Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Stat. **27**(4), 887–906 (1956)
6. Policello, G.E.: Conditional maximum likelihood estimation in gaussian mixtures. In: Taillie, C., Patil, G.P., Baldessari, B.A. (eds.) Statistical Distributions in Scientific Work. NATO Advanced study Institutes Series, vol. 79, pp. 111–125. Springer, Dordrecht (1981)

7. Ranalli, M., Lindsay, B.G., Hunter, D.R.: A classical invariance approach to the normal mixture problem. Stat. Sinica **30**, 1235–1254 (2020)
8. Tanaka, K., Takemura, A.: Strong consistency of the maximum likelihood estimator for finite mixtures of location-scale distributions when the scale parameters are exponentially small. Bernoulli **12**(6), 1003–1017 (2006)

Limit Theorems for Convex Expectations

Jonas Blessing[1] and Michael Kupper[2(✉)]

[1] Department of Mathematics, ETH Zurich, Zurich, Switzerland
jonas.blessing@math.ethz.ch
[2] Department of Mathematics and Statistics, University of Konstanz, Konstanz, Germany
kupper@uni-konstanz.de

Abstract. Based on the Chernoff approximation, we provide a general approximation result for convex monotone semigroups. Starting with a family $(I(t))_{t\geq 0}$ of operators, the semigroup is constructed as the limit $S(t)f := \lim_{n\to\infty} I(\frac{t}{n})^n f$ and is uniquely determined by the time derivative $I'(0)f$ for smooth functions. We identify explicit conditions for the generating family $(I(t))_{t\geq 0}$ that are transferred to the semigroup $(S(t))_{t\geq 0}$ and can easily be verified in applications. Furthermore, there is a structural link between Chernoff type approximations for nonlinear semigroups and law of large numbers type results for convex expectations.

Keywords: Nonlinear semigroup · Chernoff approximation · Convex expectation · Robust limit theorem

1 Introduction

In this article we summarize the main results in [3]. For further details and proofs we refer to [3]. We establish a general approximation result for convex monotone semigroups which is based on the Chernoff approximation, see [4,6]. The idea is to start with a generating family $(I(t))_{t\geq 0}$ of operators $I(t)\colon \mathrm{C_b} \to \mathrm{C_b}$ which do not form a semigroup but have, for smooth functions, the derivative

$$I'(0)f := \lim_{h\downarrow 0} \frac{I(h)f - f}{h} \in \mathrm{C_b}.$$

In order to obtain a corresponding semigroup $(S(t))_{t\geq 0}$ on $\mathrm{C_b}$ with prescribed generator $Af = I'(0)f$, we iterate the operators $(I(t))_{t\geq 0}$ over a sequence of equidistant partitions with mesh size tending to zero and define the semigroup as the limit

$$S(t)f := \lim_{n\to\infty} I\big(\tfrac{t}{n}\big)^n f \in \mathrm{C_b}. \tag{1}$$

Since the convergence in Eq. (1) is based on relative compactness, we obtain a priori only convergence for a subsequence. Following the arguments in [2,4], one can only show that the convergent subsequence can be chosen independent

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 36–43, 2024.
https://doi.org/10.1007/978-3-031-65993-5_5

of $f \in \mathrm{C_b}$ and $t \in \mathcal{T}$, where $\mathcal{T} \subset [0,\infty)$ is countable and dense. However, if the semigroup $(S(t))_{t\geq 0}$ was uniquely determined by its infinitesimal generator $Af = I'(0)f$, one could argue that the limit does not depend on the choice of the convergent subsequence and thus obtain the desired convergence in Eq. (1). This is achieved by the comparison principle in [2] for strongly continuous convex monotone semigroups.

In Sect. 4, we show that there is a structural link between Chernoff type approximations for convex monotone semigroups of the form (1) and law of large numbers (LLN) type results for convex expectations. We obtain

$$(S(1)f)(0) = \lim_{n\to\infty} \frac{1}{n}\mathcal{E}\left[nf\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)\right] \quad \text{for all } f \in \mathrm{C_b},$$

where the generator of S is given by $(Af)(x) = \mathcal{E}[\nabla f(x)\xi_1]$ for all $f \in \mathrm{C_b^1}$ and $x \in \mathbb{R}^d$. In the sublinear case, this result goes back to Peng [10].

We need the following notations. Let $\mathrm{C_b}$ be the space of all continuous functions $f\colon \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_\infty := \sup_{x\in\mathbb{R}^d} |f(x)| < \infty$. We endow $\mathrm{C_b}$ with the mixed topology between $\|\cdot\|_\infty$ and the topology of uniform convergence on compacts sets. It is well-known, see [7, Proposition A.4], that a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{C_b}$ converges to $f \in \mathrm{C_b}$ w.r.t. the mixed topology if and only if

$$\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty \quad \text{and} \quad \lim_{n\to\infty} \|f - f_n\|_{\infty,K} = 0$$

for all compact subsets $K \subset \mathbb{R}^d$, where $\|f\|_{\infty,K} := \sup_{x\in K} |f(x)|$. Subsequently, all limits in $\mathrm{C_b}$ are understood w.r.t. the mixed topology and compact subsets are denoted by $K \Subset \mathbb{R}^d$. For more details, we refer to [7] and the references therein. Moreover, the set of all real-valued functions is endowed with the pointwise order, i.e., for functions $f, g\colon \mathbb{R}^d \to \mathbb{R}$ we write $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in \mathbb{R}^d$. Let

$$B_{\mathrm{C_b}}(r) := \{f \in \mathrm{C_b}\colon \|f\|_\infty \leq r\} \quad \text{and} \quad B_{\mathbb{R}^d}(r) := \{x \in \mathbb{R}^d\colon |x| \leq r\}$$

be the closed balls with radius $r \geq 0$ around zero. Let $\mathrm{Lip_b}$ be the space of all bounded Lipschitz continuous functions $f\colon \mathbb{R}^d \to \mathbb{R}$ and, for every $r \geq 0$, the set $\mathrm{Lip_b}(r)$ consists of all bounded r-Lipschitz functions $f\colon \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_\infty \leq r$. Let $(\tau_x f)(y) := f(x+y)$ for all $x, y \in \mathbb{R}^d$ and $f\colon \mathbb{R}^d \to \mathbb{R}$ be the shift operators.

2 Chernoff Approximation

We discuss the Chernoff approximation of convex monotone semigroups.

Definition 1. *A family $(S(t))_{t\geq 0}$ of operators $S(t)\colon \mathrm{C_b} \to \mathrm{C_b}$ is called a semigroup if $S(0) = \mathrm{id}_{\mathrm{C_b}}$ and $S(s+t)f = S(s)S(t)f$ for all $s, t \geq 0$ and $f \in \mathrm{C_b}$. The semigroup is convex and monotone if the mapping $\mathrm{C_b} \to \mathbb{R}$, $f \mapsto (S(t)f)(x)$ is convex and monotone for all $t \geq 0$ and $x \in \mathbb{R}^d$. Moreover, the semigroup is*

called strongly continuous if the mapping $\mathbb{R}_+ \to \mathrm{C_b},\ t \mapsto S(t)f$ *is continuous for all* $f \in \mathrm{C_b}$*. The generator is defined by*

$$A\colon D(A) \to \mathrm{C_b},\ f \mapsto \lim_{h \downarrow 0} \frac{S(h)f - f}{h},$$

where the domain $D(A)$ *consists of all* $f \in \mathrm{C_b}$ *such that the previous limit exists.*

Let $(I(t))_{t \geq 0}$ be a family of operators $I(t)\colon \mathrm{C_b} \to \mathrm{C_b}$ and $(h_n)_{n \in \mathbb{N}} \subset (0,1]$ be a sequence with $h_n \to 0$. For every $t \geq 0$, $n \in \mathbb{N}$ and $f \in \mathrm{C_b}$, we define

$$I(\pi_n^t)f := I(h_n)^{k_n^t} f = \underbrace{\big(I(h_n) \circ \ldots \circ I(h_n)\big)}_{k_n^t \text{ times}} f,$$

where $k_n^t := \max\{k \in \mathbb{N}_0\colon kh_n \leq t\}$ and $\pi_n^t := \{kh_n \wedge t\colon k \in \mathbb{N}_0\}$ denotes the equidistant partition of $[0,t]$ with mesh size h_n. We assume the following conditions:

(i) $I(0) = \mathrm{id}_{\mathrm{C_b}}$.
(ii) $I(t)$ is convex and monotone with $I(t)0 = 0$ for all $t \geq 0$.
(iii) There exists $\omega \geq 0$ with

$$\|I(t)f - I(t)g\|_\infty \leq e^{\omega t}\|f - g\|_\infty \quad \text{for all } t \in [0,1] \text{ and } f, g \in \mathrm{C_b}.$$

(iv) There exist $t_0 > 0$, $\delta \in (0,1]$ and $L \geq 0$ with

$$\|I(t)(\tau_x f) - \tau_x I(t)f\|_\infty \leq Lrt|x|$$

for all $t \in [0,t_0]$, $x \in B_{\mathbb{R}^d}(\delta)$, $r \geq 0$ and $f \in \mathrm{Lip_b}(r)$.
(v) It holds that $\mathrm{C_b^\infty} \subset \mathcal{L}^I$ and $I'(0)f := \lim_{h \downarrow 0} \frac{I(h)f - f}{h} \in \mathrm{C_b}$ exists for all $f \in \mathrm{C_b^\infty}$. Here, $\mathrm{C_b^\infty}$ is the space of all infinitely differentiable functions $f \in \mathrm{C_b}$ such that all partial derivatives are in $\mathrm{C_b}$, and $\mathcal{L}^I$ denotes the Lipschitz set of all $f \in \mathrm{C_b}$ for which there exist $c \geq 0$ and $t_0 > 0$ with $\|I(t)f - f\|_\infty \leq ct$ for all $t \in [0,t_0]$.
(vi) For every $\epsilon > 0$, $r, T \geq 0$ and $K \Subset \mathbb{R}^d$, there exist $c \geq 0$ and $K' \Subset \mathbb{R}^d$ with

$$\|I(\pi_n^t)f - I(\pi_n^t)g\|_{\infty,K} \leq \|f - g\|_{\infty,K'} + \epsilon$$

for all $t \in [0,T]$, $f, g \in B_{\mathrm{C_b}}(r)$ and $n \in \mathbb{N}$.
(vii) It holds that $I(t)\colon \mathrm{Lip_b}(r) \to \mathrm{Lip_b}(e^{\omega t}r)$ for all $r, t \geq 0$.

The key idea here is to identify conditions on the one-step operators $I(t)$ that are preserved during the iteration and thus transfer to the associated semigroup $(S(t))_{t \geq 0}$. Although condition (vi) is stated for the iterated operators $I(\pi_n^t)$, there are several sufficient conditions for the one-step operators $I(t)$ that can be verified in applications. Such conditions are more systematically studied in [5, Sect. 2.5]. In many applications most of the previous conditions can easily be verified while the computation of the derivative $I'(0)f$ usually takes some computational effort. However, for smooth functions this can often be achieved by elementary arguments such as Taylor's formula with a suitable remainder estimate.

Theorem 1. *Let $(I(t))_{t\geq 0}$ be a family of operators which satisfy the previous conditions (i)–(vii). Then, there exists a strongly continuous convex monotone semigroup $(S(t))_{t\geq 0}$ on $\mathrm{C_b}$ with*

$$S(t)f = \lim_{n\to\infty} I(\pi_n^t)f \quad \text{for all } t \geq 0 \text{ and } f \in \mathrm{C_b}.$$

Furthermore, the semigroup has the following properties:

(i) It holds that $f \in D(A)$ and $Af = I'(0)f$ for all $f \in \mathrm{C_b}$ such that $I'(0)f \in \mathrm{C_b}$ exists. In particular, this is valid for all $f \in \mathrm{C_b^\infty}$.
(ii) It holds that $\|S(t)f - S(t)g\|_\kappa \leq e^{\omega t}\|f-g\|_\infty$ for all $t \geq 0$ and $f, g \in \mathrm{C_b}$.
(iii) For every $\epsilon > 0$, $r, T \geq 0$ and $K \Subset \mathbb{R}^d$, there exist $K' \Subset \mathbb{R}^d$ and $c \geq 0$ with

$$\|S(t)f - S(t)g\|_{\infty,K} \leq c\|f-g\|_{\infty,K'} + \epsilon$$

for all $t \in [0,T]$ and $f, g \in B_{\mathrm{C_b}}(r)$.

3 Convex Expectation Spaces

Nonlinear expectations were introduced by Peng, see [11] for a detailed discussion, and are closely related to several other concepts. For instance, sublinear expectations are called upper expectation in robust statistics [8], upper coherent prevision in the theory of imprecise probabilities [12] and coherent risk measure [1] in mathematical finance. All inequalities in the following definition are understood pointwise.

Definition 2. *Let Ω be a set and $\mathcal{H}$ a linear space of functions $X\colon \Omega \to \mathbb{R}$ with $c \in \mathcal{H}$ and $|X| \in \mathcal{H}$ for all $c \in \mathbb{R}$ and $X \in \mathcal{H}$. A convex expectation is a functional $\mathcal{E}\colon \mathcal{H} \to \mathbb{R}$ with*

- *$\mathcal{E}[c] = c$ for all $c \in \mathbb{R}$,*
- *$\mathcal{E}[X] \leq \mathcal{E}[Y]$ for all $X, Y \in \mathcal{H}$ with $X \leq Y$,*
- *$\mathcal{E}[\lambda X + (1-\lambda)Y] \leq \lambda\mathcal{E}[X] + (1-\lambda)\mathcal{E}[Y]$ for all $X, Y \in \mathcal{H}$ and $\lambda \in [0,1]$.*

The triplet $(\Omega, \mathcal{H}, \mathcal{E})$ is called a convex expectation space.

Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a convex expectation space with $f(X_1, \ldots, X_n) \in \mathcal{H}$ for all $n \in \mathbb{N}$, $f \in \mathrm{C_b}(\mathbb{R}^n)$ and $X \in \mathcal{H}^n$. Let $m, n \in \mathbb{N}$, $X \in \mathcal{H}^m$ and $Y \in \mathcal{H}^n$. We say that X and Y are identically distributed if $m = n$ and

$$\mathcal{E}[f(X)] = \mathcal{E}[f(Y)] \quad \text{for all } f \in \mathrm{Lip_b}(\mathbb{R}^m).$$

We say that Y is independent of X if

$$\mathcal{E}[f(X,Y)] = \mathcal{E}\left[\mathcal{E}[f(x,Y)]\big|_{x=X}\right] \quad \text{for all } f \in \mathrm{Lip_b}(\mathbb{R}^m \times \mathbb{R}^n).$$

In general, the statements "Y is independent of X" and "X is independent of Y" are not equivalent, see [11, Example 1.3.15] for a counterexample.

Definition 3. *Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a convex expectation space and $(X_n)_{n\in\mathbb{N}} \subset \mathcal{H}^d$ be a sequence of random vectors for some $d \in \mathbb{N}$.*

(i) We say that $(X_n)_{n\in\mathbb{N}}$ is independent and identically distributed (iid) if X_m and X_n have the same distribution and X_{n+1} is independent of $(X_1, \dots, X_n)$ for all $m, n \in \mathbb{N}$.

(ii) We say that $(X_n)_{n\in\mathbb{N}}$ converges in distribution if the sequence $(\mathcal{E}[f(X_n)])_{n\in\mathbb{N}}$ converges for all $f \in \mathrm{Lip}_\mathrm{b}(\mathbb{R}^d)$.

4 First Order Scaling Limits

Let $(\xi_n)_{n\in\mathbb{N}} \subset \mathcal{H}^d$ be an iid sequence of random vectors on a convex expectation space $(\Omega, \mathcal{H}, \bar{\mathcal{E}})$ and $(\psi_n)_{n\in\mathbb{N}}$ be a sequence of recursively defined functions $\psi_n \colon \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d$. We define a sequence $(X_n)_{n\in\mathbb{N}}$ of random vectors by

$$X_n := \psi_n\left(\tfrac{1}{n}, 0, \xi_1, \dots, \xi_n\right)$$

whose behaviour in the limit we are interested in. We define $I(0) := \mathrm{id}_{\mathrm{C_b}}$ and

$$(I(t)f)(x) := t\bar{\mathcal{E}}\left[\frac{1}{t} f\big(\psi_1(t, x, \xi_1)\big)\right]$$

for all $t > 0$, $f \in \mathrm{C_b}$ and $x \in \mathbb{R}^d$. Then, the independence of $(\xi_n)_{n\in\mathbb{N}}$ implies

$$\big(I\big(\tfrac{1}{n}\big)^n f\big)(0) = \frac{1}{n}\bar{\mathcal{E}}\left[nf(X_n)\right].$$

Hence, the operator theoretic result from Sect. 2 can be formulated equivalently by using classic probabilistic notation. In Sect. 4.1, we prove the previously discussed main result for first order scaling limits as well as some immediate consequences.

4.1 LLN for Convex Expectations and Large Deviations

Let $\mathrm{C_{lin}}$ be the space of all continuous functions $f \colon \mathbb{R}^d \to \mathbb{R}$ with linear growth $\sup_{x\in\mathbb{R}^d} |f(x)(1 + |x|)^{-1}| < \infty$. Let $\mathcal{E} \colon \mathrm{C_{lin}} \to \mathbb{R}$ be a convex expectation which is continuous from above, i.e., $\mathcal{E}[f_n] \downarrow 0$ for all $(f_n)_{n\in\mathbb{N}} \subset \mathrm{C_{lin}}$ with $f_n \downarrow 0$.

Let $\psi \colon \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function which satisfies the following conditions:

(i) There exists $L \geq 0$ such that, for all $t \in [0, 1]$ and $x, y, z \in \mathbb{R}^d$,

$$|x + \psi(t, y, z) - \psi(t, x + y, z)| \leq Lt|x|.$$

Furthermore, it holds that $|\psi(t, x, y) - x| \leq L(1 + |y|)t$ for all $t \in [0, 1]$ and $x, y \in \mathbb{R}^d$.

(ii) There exists a continuous function $\psi_0 \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ with

$$\lim_{h \downarrow 0} \sup_{x,y \in K} \left| \frac{\psi(h,x,y) - x}{h} - \psi_0(x,y) \right| = 0 \quad \text{for all } K \Subset \mathbb{R}^d.$$

The previous conditions are satisfied for $\psi(t,x,y) := x + ty$ and $\psi_0(x,y) := y$ in which case it holds $\psi_n(\frac{1}{n}, x, y_1, \dots, y_n) = x + \frac{1}{n}\sum_{i=1}^n y_i$. Another valid choice would be $\psi(t,x,y) := x + t\varphi(x) + ty$ and $\psi_0(x,y) := \varphi(x) + y$ for some Lipschitz continuous function $\varphi \colon \mathbb{R}^d \to \mathbb{R}^d$ describing a perturbation of the sample ξ_{n+1} by a nonlinear function depending on the average of previous samples.

We define $I(0) := \mathrm{id}_{\mathrm{C_b}}$ and, for every $t > 0$, $f \in \mathrm{C_b}$ and $x \in \mathbb{R}^d$,

$$(I(t)f)(x) := t\mathcal{E}\left[\frac{1}{t} f\big(\psi(t,x,\cdot)\big)\right]$$

Let $\mathrm{C_b^1}$ consist of all differentiable functions $f \in \mathrm{C_b}$ such that all partial derivatives are in $\mathrm{C_b}$ and denote by $xy := \langle x, y \rangle$ the Euclidean inner product on $\mathbb{R}^d$.

Theorem 2. *There exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $\mathrm{C_b}$ with*

$$S(t)f = \lim_{n \to \infty} I\big(\tfrac{t}{n}\big)^n f \in \mathrm{C_b} \quad \textit{for all } t \geq 0 \textit{ and } f \in \mathrm{C_b}$$

which has the properties (i)-(iii) from Theorem 1 with $\omega = 0$. For every $f \in \mathrm{C_b^1}$, it holds that $f \in D(A)$ and

$$(Af)(x) = \mathcal{E}[\nabla f(x) \psi_0(x, \cdot)] \quad \textit{for all } x \in \mathbb{R}^d.$$

For every convex expectation space $(\Omega, \mathcal{H}, \bar{\mathcal{E}})$ and iid sequence $(\xi_n)_{n \in \mathbb{N}} \subset \mathcal{H}^d$ with $\bar{\mathcal{E}}[f(\xi_n)] = \mathcal{E}[f]$ for all $f \in \mathrm{C_b}$,

$$(S(1)f)(0) = \lim_{n \to \infty} \frac{1}{n} \bar{\mathcal{E}}\left[n f(X_n)\right] \quad \textit{for all } f \in \mathrm{C_b},$$

where $X_n := \psi_n\big(\frac{1}{n}, 0, \xi_1, \dots, \xi_n\big)$ and $\psi_n \colon \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d$ is recursively defined by $\psi_1 := \psi$ and $\psi_{n+1}(t,x,y_1,\dots,y_{n+1}) := \psi\big(t, \psi_n(t,x,y_1,\dots,y_n), y_{n+1}\big)$.

To give an explicit example for the sequence $(\psi_n)_{n \in \mathbb{N}}$, we consider the case that the sample ξ_{n+1} is randomly shifted by a nonlinear function depending on the average of the previous samples $\xi_1, \dots, \xi_n$.

Corollary 1. *Let $\psi(t,x,y) := x + \varphi(t,x) + ty$ for a function $\varphi \colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ such that there exist $L \geq 0$ with $\varphi(t,\cdot) \in \mathrm{Lip_b}(Lt)$ for all $t \geq 0$. Furthermore, we assume that the limit $\varphi_0 := \lim_{h \downarrow 0} \frac{\varphi(h,\cdot)}{h} \in \mathrm{C_b}$ exists. Then, there exists a strongly continuous convex monotone semigroup $(S(t))_{t \geq 0}$ on $\mathrm{C_b}$ with*

$$S(t)f = \lim_{n \to \infty} I\big(\tfrac{t}{n}\big)^n f \in \mathrm{C_b} \quad \textit{for all } t \geq 0 \textit{ and } f \in \mathrm{C_b}$$

which has the properties (i)-(iii) from Theorem 1 with $\omega = 0$*. For every* $f \in \mathrm{C}_\mathrm{b}^1$ *and* $x \in \mathbb{R}^d$*, it holds that* $f \in D(A)$ *and*

$$(Af)(x) = \mathcal{E}[\nabla f(x)\xi] + \varphi_0(x)\nabla f(x), \quad \text{where } \xi := \mathrm{id}_{\mathbb{R}^d}.$$

For every convex expectation space $(\Omega, \mathcal{H}, \bar{\mathcal{E}})$ *and iid sequence* $(\xi_n)_{n\in\mathbb{N}} \subset \mathcal{H}^d$ *with* $\bar{\mathcal{E}}[f(\xi_n)] = \mathcal{E}[f]$ *for all* $f \in \mathrm{C}_\mathrm{b}$*,*

$$(S(1)f)(0) = \lim_{n\to\infty} \frac{1}{n}\bar{\mathcal{E}}\left[nf(X_n)\right] \quad \text{for all } f \in \mathrm{C}_\mathrm{b},$$

where $X_n := \psi_n(\frac{1}{n}, 0, \xi_1, \ldots, \xi_n)$*.*

For non-perturbed averaged sums of iid samples, the semigroup can be represented explicitly by the Hopf–Lax formula.

Theorem 3. *Let* $\psi(t, x, y) := x + ty$ *for all* $t \geq 0$ *and* $x, y \in \mathbb{R}^d$*. Then, the semigroup* $(S(t))_{t\geq 0}$ *from Theorem 2 has the representation*

$$(S(t)f)(x) = \sup_{y\in\mathbb{R}^d} \big(f(x+ty) - \varphi(y)t\big) \quad \text{for all } t \geq 0, f \in \mathrm{C}_\mathrm{b} \text{ and } x \in \mathbb{R}^d,$$

where $\varphi(y) := \sup_{z\in\mathbb{R}^d}(yz - \mathcal{E}[z\xi])$ *and* $\xi := \mathrm{id}_{\mathbb{R}^d}$*. For every* $f \in \mathrm{C}_\mathrm{b}^1$*, it holds that*

$$(Af)(x) = \mathcal{E}[\nabla f(x)\xi] \quad \text{for all } x \in \mathbb{R}^d.$$

For every convex expectation space $(\Omega, \mathcal{H}, \bar{\mathcal{E}})$ *and iid sequence* $(\xi_n)_{n\in\mathbb{N}} \subset \mathcal{H}^d$ *with* $\bar{\mathcal{E}}[f(\xi_n)] = \mathcal{E}[f]$ *for all* $f \in \mathrm{C}_\mathrm{b}$*,*

$$\lim_{n\to\infty} \frac{1}{n}\bar{\mathcal{E}}\left[nf\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)\right] = \sup_{x\in\mathbb{R}^d} \big(f(x) - \varphi(x)\big) \quad \text{for all } f \in \mathrm{C}_\mathrm{b}.$$

4.2 Applications

As a result of Theorem 3, we obtain Cramér's theorem as LLN for the entropic risk measure. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\xi_n)_{n\in\mathbb{N}}$ be an iid sequence of random vectors $\xi_n \colon \Omega \to \mathbb{R}^d$ with $\mathbb{E}[e^{c|\xi_1|}] < \infty$ for all $c \geq 0$. Let

$$\Lambda \colon \mathbb{R}^d \to \mathbb{R},\ x \mapsto \log\big(\mathbb{E}[e^{x\xi_1}]\big)$$

be the logarithmic moment generating function and denote by $\Lambda^* \colon \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \sup_{y\in\mathbb{R}^d}\big(xy - \Lambda(y)\big)$ its convex conjugate. Moreover, we define $X_n := \frac{1}{n}\sum_{i=1}^n \xi_i$ for all $n \in \mathbb{N}$.

Corollary 2. *For every* $f \in \mathrm{C}_\mathrm{b}$*,*

$$\lim_{n\to\infty} \frac{1}{n}\log\big(\mathbb{E}\big[e^{nf(X_n)}\big]\big) = \sup_{x\in\mathbb{R}^d} \big(f(x) - \Lambda^*(x)\big).$$

Furthermore, if we only require that $(\xi_n)_{n\in\mathbb{N}}$ has finite p-th moments instead of finite exponential moments, we still obtain polynomial convergence rates as in [9]. For the following theorem, let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\xi_n)_{n\in\mathbb{N}}$ be an iid sequence of random vectors $\xi_n\colon \Omega \to \mathbb{R}^d$. Furthermore, let $\varphi\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be a function such that there exist $L \geq 0$ with $\varphi(t,\cdot) \in \mathrm{Lip}_{\mathrm{b}}(Lt)$ for all $t \geq 0$ and suppose that the limit $\varphi_0 := \lim_{h\downarrow 0} \frac{\varphi(h,\cdot)}{h} \in \mathrm{C}_{\mathrm{b}}$ exists. We define $\psi(t,x,y) := x + \varphi(t,x) + ty$ and recursively a sequence $(\psi_n)_{n\in\mathbb{N}}$ of functions $\psi_n\colon \mathbb{R}_+ \times \mathbb{R}^d \times (\mathbb{R}^d)^n \to \mathbb{R}^d$ by $\psi_1 := \psi$ and

$$\psi_{n+1}(t,x,y_1,\dots,y_{n+1}) := \psi\big(t,\psi_n(t,x,y_1,\dots,y_n),y_{n+1}\big).$$

Finally, let $X_n := \psi_n(\frac{1}{n},0,\xi_1,\dots,\xi_n)$ for all $n \in \mathbb{N}$.

Theorem 4. *Assume that* $\mathbb{E}[|\xi_1|^p] < \infty$ *for some* $p \in (1,\infty)$. *Define*

$$\Lambda\colon \mathbb{R}^d \to \mathbb{R},\ x \mapsto \inf\{m \in \mathbb{R}\colon \mathbb{E}[((1 + x\xi_1 - m)^+)^p] \leq 1\}$$

and the convex conjugate $\Lambda^*\colon \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \sup_{y\in\mathbb{R}^d}(xy - \Lambda(y))$. *Then, for every closed subset* $A \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} n^{p-1}\mathcal{P}(X_n \in A) \leq \left(\inf_{x\in A_L} \Lambda^*(x)\right)^{-p},$$

where $A_L := \{x + y\colon x \in A,\ y \in B_{\mathbb{R}^d}(L)\}$.

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
2. Blessing, J., Denk, R., Kupper, M., Nendel, M.: Convex monotone semigroups and their generators with respect to Γ-convergence. Preprint arXiv:2202.08653 (2022)
3. Blessing, J., Kupper, M.: Nonlinear semigroups and limit theorems for convex expectations. Preprint arXiv:2210.14096 (2022)
4. Blessing, J., Kupper, M.: Nonlinear semigroups built on generating families and their Lipschitz sets. Potential Anal. **59**(3), 857–895 (2023)
5. Blessing, J., Kupper, M., Nendel, M.: Convergence of infinitesimal generators and stability of convex monotone semigroups. Preprint arXiv:2305.18981 (2023)
6. Chernoff, P.R.: Note on product formulas for operator semigroups. J. Funct. Anal. **2**, 238–242 (1968)
7. Goldys, B., Nendel, M., Röckner, M.: Operator semigroups in the mixed topology and the infinitesimal description of Markov processes. Preprint arXiv:2204.07484 (2022)
8. Huber, P.J.: Robust statistics. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1981)
9. Lacker, D.: A non-exponential extension of Sanov's theorem via convex duality. Adv. Appl. Probab. **52**(1), 61–101 (2020)
10. Peng, S.: A new central limit theorem under sublinear expectations. Preprint arXiv:0803.2656 (2008)
11. Peng, S.: Nonlinear expectations and stochastic calculus under uncertainty. Probability Theory and Stochastic Modelling, Springer, Berlin (2019)
12. Walley, P.: Statistical reasoning with imprecise probabilities. Monographs on Statistics and Applied Probability. Chapman and Hall Ltd, London (1991)

The Joint Weighted Average (JWA) Operator

Stephen B. Broomell[1] and Christian Wagner[2(✉)]

[1] GRID—Global Risk and Individual Decisions Laboratory, Purdue University, West Lafayette, IN 47907, USA
Broomell@purdue.edu

[2] University of Nottingham, LUCID—Lab for Uncertainty in Data and Decision Making, Nottingham, UK
Christian.Wagner@nottingham.ac.uk

Abstract. Information aggregation is a vital tool for human and machine decision making in the presence of uncertainty. Traditionally, approaches to aggregation broadly diverge into two categories, those which attribute a worth or weight to information sources and those which attribute said worth to the evidence arising from said sources. The latter is pervasive in the physical sciences, underpinning linear order statistics and enabling non-linear aggregation. The former is popular in the social sciences, providing interpretable insight on the sources. While prior work has identified the need to apply both approaches simultaneously, it has yet to conceptually integrate both approaches and provide a semantic interpretation of the arising aggregation approach. Here, we conceptually integrate both approaches in a novel joint weighted averaging operator. We leverage compositional geometry to underpin this integration, showing how it provides a systematic basis for the combination of weighted aggregation operators–which has thus far not been considered in the literature. We proceed to show how the resulting operator systematically integrates a priori beliefs about the worth of both sources and evidence, reflecting the semantic integration of both weighting strategies. We conclude and highlight the potential of the operator across disciplines, from machine learning to psychology.

Keywords: Decision making · Aggregation · OWA · Fusion · Composition · Weighting · Evidence

1 Introduction

Whether it be for a computerized system or for human decision making, evidence derived from information sources serves as decision inputs, and the discord between pieces of evidence can often be addressed by considering multiple of said sources. For example, museums commonly use multiple thermometers to ensure the maintenance of an ideal temperature reliably, while expert committees rely on the presence of multiple experts to address discord and individual knowledge-gaps.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 44–58, 2024.
https://doi.org/10.1007/978-3-031-65993-5_6

Two different approaches are commonly used to aggregate information. The first approach is a linearly weighted average (LWA), which underlies many approaches to information aggregation for wisdom of crowds [2,5,6,8]. The second approach is operationalized by ordered combinations of the evidence or judgment arising from the information sources. Flexible non-linear operators such as linear order-statistics have been applied with great success in applications from cyber-security to data-fusion [12,13]. For example, the Order Weighted Average (OWA), [20] retains much of the simplicity and interpretability of the LWA but allows for non-linear combinations of evidence. To achieve this, the OWA focuses *not* on the notion of worth arising from each source, but instead on the worth associated with the actual information, i.e. the *evidence* arising from each given source at a given point in time.

Prior literature has developed operators that combine multiple operators, such as multiple OWA operators [11]. We focus here on operators that attempt to combine the advantages of the OWA and the LWA. These include the Weighted OWA (WOWA) [16,17] the Hybrid Weighted Average (HWA) [19], the Ordered Weighted Averaging – Weighted Averaging (OWAWA) [10], and the Standard Deviation OWA (SDOWA) [4]. The former two operators have limitations to their behavior, as outlined by Llamazares [9], and the latter two operators are based on the partial consideration of the OWA and LWA (i.e. each is used to a different degree), rather than the *joint* consideration of them. We show that all prior operators lack a mathematical framework that provides a semantic justification and definition for how the LWA and OWA should be combined.

In this paper, we introduce such a mathematical framework for creating joint weights and explore the nature of a joint focus on source-and-evidence specific worth of information. In Sect. 2 we define each of the extant aggregation operators. In Sect. 3 we discuss the assumptions and properties of information sources and evidence that would lead a decision maker to apply LWA, OWA, or some combination of the two for their information aggregation needs, showing inconsistent behavior in each of the existing operators. Next, we introduce compositional geometry [1,15] and discuss how this mathematical framework accounts for prior inconsistencies, defining precisely joint weights that create aggregates based on both source and evidence. In Sect. 4 we introduce our new aggregation operator–the Joint Weighted Average (JWA)–and in Sect. 5 we provide an illustrative experiment. Finally, Sect. 6 is the discussion and conclusion.

2 Preliminaries

2.1 Basic Aggregation Operators and Their Properties

Prior work has established properties of aggregation functions that define their behavior (see [4,9]). Some of these properties are defined because they are desirable, but some are defined in order to establish logical bounds on the behavior of an aggregator. Let $F(\boldsymbol{x})$ be a function that takes in a vector $\boldsymbol{x} \in R^n$ of observations and generates an aggregate. The function $F(\boldsymbol{x})$ may or may not have the following properties:

1. **Compensativeness.** For all $\boldsymbol{x} \in R^n$, $min(\boldsymbol{x}) \leq F(\boldsymbol{x}) \leq max(\boldsymbol{x})$.
2. **Idempotency.** For all $x \in R$, $F(x, \ldots, x) = x$.
3. **Homogeneity of degree 1.** For all $\boldsymbol{x} \in R^n$ and $\lambda > 0$, $F(\lambda \boldsymbol{x}) = \lambda F(\boldsymbol{x})$.
4. **Monotonicity.** For all $\boldsymbol{x}, \boldsymbol{y} \in R^n$, if $x_i \geq y_i$ for all $i \in \{1, \ldots, n\}$ then $F(\boldsymbol{x}) \geq F(\boldsymbol{y})$.
5. **Commutativity.** For all $\boldsymbol{x} \in R^n$ and for all permutations of $\boldsymbol{x}$ given by $\boldsymbol{x}_\pi$, $F(\boldsymbol{x}_\pi) = F(\boldsymbol{x})$.

The Linear Weighted Average (LWA) Let $\boldsymbol{w}$ be a set of weights such that $\boldsymbol{w} \in [0, 1]^n$ and $\sum_{i=1}^{n} w_i = 1$. The LWA is a linear combination of the evidence provided by each information source (x_i) given by[1],

$$LWA(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i. \tag{1}$$

The LWA satisfies properties 1–4, and conditionally satisfies property 5 if and only if $w_i = 1/n$ for all $i \in \{1, \ldots, n\}$.

The Order Weighted Average (OWA) [20] Let $\boldsymbol{v}$ be a vector of weights such that $\boldsymbol{v} \in [0, 1]^n$ and $\sum_{i=1}^{n} v_i = 1$. Let $\pi()$ be a permutation function such that $x_{\pi(1)} \geq ... \geq x_{\pi(n)}$. The OWA is an ordered combination of the evidence provided by each information source (x_i) given by,

$$OWA(\boldsymbol{x}) = \sum_{i=1}^{n} v_i x_{\pi(i)} \tag{2}$$

The OWA satisfies properties 1–5.

2.2 Combination Aggregation Operators and Their Properties

Aggregation operators that combine the LWA and OWA can introduce additional complexity. The behavior of such combined operators $F'(\boldsymbol{x})$ has been understood in prior work by an additional set of properties that define their relationship to the LWA and OWA as outlined by [4,9,16].

6. **Collapsing.**[2] For all $\boldsymbol{x} \in R^n$, if $w_1 = \ldots = w_n$ then, $F'(\boldsymbol{x}) = OWA(\boldsymbol{x})$ and if $v_1 = \ldots = v_n$ then $F'(\boldsymbol{x}) = LWA(\boldsymbol{x})$.
7. **Internal Boundedness.** For all $\boldsymbol{x} \in R^n$, $min(OWA(\boldsymbol{x}), LWA(\boldsymbol{x})) \leq F'(\boldsymbol{x}) \leq max(OWA(\boldsymbol{x}), LWA(\boldsymbol{x}))$.
8. **Transitivity of Equality.** For all $\boldsymbol{x} \in R^n$, if $OWA(\boldsymbol{x}) = LWA(\boldsymbol{x}) = k$ then $F'(\boldsymbol{x}) = k$.

[1] We note that the indices link the given source and associated weight.

[2] This property captures whether a combined operator collapses to one of its constituent operators under a specific configuration of the weight parameters.

The Weighted OWA (WOWA) [16] Let $\boldsymbol{w}$ be the vector of weights defined by LWA and $\boldsymbol{v}$ be the vector of weights defined by OWA. Let $\pi()$ be the permutation function defined by OWA. Let the function s() interpolate between points $\left(i/n, \sum_{j=1}^{i} v_j\right)$ together with $(0,0)$. A new weight vector $\boldsymbol{\omega}$ is computed as,

$$\omega_i = s\left(\sum_{j=1}^{i} w_{\pi(j)}\right) - s\left(\sum_{j=1}^{i-1} w_{\pi(j)}\right). \tag{3}$$

The WOWA is an ordered combination of the evidence provided by each information source (x_i) given by,

$$WOWA(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i x_{\pi(i)}. \tag{4}$$

The WOWA satisfies properties 1–4 and 6 [9]. It conditionally satisfies property 5 if and only if $w_i = 1/n$ for all $i \in \{1, \ldots, n\}$ and it collapses to the OWA.

The Hybrid Weighted Average (HWA) [19] Let $\boldsymbol{w}$ be the vector of weights defined by LWA and $\boldsymbol{v}$ be the vector of weights defined by OWA. Let $\pi()$ be a permutation function such that $nw_{\pi(1)}x_{\pi(1)} \geq \ldots \geq nw_{\pi(n)}x_{\pi(n)}$. The HWA is an ordered combination of the evidence provided by each information source (x_i) given by,

$$HWA(\boldsymbol{x}) = \sum_{i=1}^{n} v_i (nw_{\pi(i)}x_{\pi(i)}). \tag{5}$$

The HWA satisfies properties 3, 4, and 6 [9]. It conditionally satisfies properties 1 and 2 if and only if it collapses to the LWA or OWA, and property 5 if and only if it collapses to OWA.

Ordered Weighted Average—Weighted Average (OWAWA) [10] A simple method for combining the LWA and OWA is to generate a weighted combination of the two aggregators. The OWAWA operator is given by,

$$OWAWA(\boldsymbol{x}) = \alpha LWA(\boldsymbol{x}) + (1-\alpha)OWA(\boldsymbol{x}). \tag{6}$$

The parameter $\alpha \in [0,1]$ defines the relative contribution of the LWA and OWA to the final aggregate. The OWAWA satisfies properties 1–4, 7, and 8 [4]. It conditionally satisfies property 5 if and only if $w_i = 1/n$ for all $i \in \{1, \ldots, n\}$ or $\alpha = 0$.

Standard Deviation OWA (SDOWA) [4] Let the function $sd()$ return the standard deviation of a vector. Let the function $G()$ be defined as,

$$G(\boldsymbol{w}, \boldsymbol{v}) = \frac{sd(\boldsymbol{w})}{sd(\boldsymbol{w}) + sd(\boldsymbol{v})}. \tag{7}$$

The SDOWA operator is given by,

$$\begin{aligned} SDOWA(\boldsymbol{x}) = {} & G(\boldsymbol{w}, \boldsymbol{v}) LWA(\boldsymbol{x}) \\ & + (1 - G(\boldsymbol{w}, \boldsymbol{v}))OWA(\boldsymbol{x}). \end{aligned} \tag{8}$$

The SDOWA satisfies properties 1–4 and 6–8 [4]. It conditionally satisfies property 5 if and only if $w_i = 1/n$ for all $i \in \{1, \ldots, n\}$.

3 Theoretical Context

In order to create an operator that fully integrates the LWA and the OWA operators, it is valuable to take a step back and conceptually understand the properties of the evidence we wish to aggregate and the desired outcome of integrating two aggregators. We start with a conceptual example where linear and order weighting both apply, and show how each aggregator defined above fails to produce a coherent aggregate.

Consider the problem of estimating how many product items to produce, using judgments from 3 experts. First, we may have some prior knowledge of the relative performance of each expert such that we assign weights {0.60, 0.30, 0.10} to experts 1, 2 and 3 respectively. These weights represent an LWA, the goals of which are to give weight to the experts (sources) that matches their relative reliability or validity. Second, we may have a desire to avoid underproduction of our product such that we assign ordered weights {0.45, 0.50, 0.05} to the highest, middle, and lowest judgment produced by the experts. These weights represent an OWA, the goals of which are to consider the middle judgment most strongly along with the highest judgment, and put very little weight on the lowest judgment.

It is clear that beyond the linear and non-linear characteristics of the LWA and OWA, both operators operationalize fundamentally different foci, with the LWA *focusing* on the *worth* of the information sources, and the OWA focusing on the *worth* of the evidence or judgment contributed by the sources. Another way of looking at this is that the LWA operationalizes a strategy aligned with the sources, for example here: combine considering the expertise of the experts; while the OWA operationalizes a strategy aligned with the evidence, for example here: avoid under-production. If we apply the OWA (or the LWA) in isolation, we implicitly assume that all experts (or evidence orderings) are equally valid–we apply aggregation based on one *or* the other strategy–not both. However, we wish to apply both sets of weights to aggregate the expert judgments to account for both strategies.

Table 1 displays aggregation results for each of the operators defined in the previous section using example evidence. We display the weight that was assigned to each judgement and the final aggregates of each operator. The weights show how well each operator represents the strategies defined by the LWA and OWA weights. Similar to examples provided by [9], the WOWA exhibits strange behavior by shifting some weight from expert 2 to expert 3 (the least valued expert and order), appearing to act against the goals of both the LWA and the OWA weights. The HWA generates weights for the evidence that do not sum to one, as such, the aggregate is outside the bounds of the data, violating compensativeness. However, the HWA weights do appear to represent both the LWA and OWA strategies, reducing the weight for expert 3's judgment (least trusted expert and

order) and increasing the weight of expert 1 (most trusted expert and good order location).

We can also illustrate the difference between a weighted combination of both operators and the joint consideration of both aggregation strategies. Based on the results in Table 1, any weighted aggregation of LWA and OWA (which includes the SDOWA and OWAWA) would by definition generate an estimate in the range [66, 70], i.e. between the output produced by the OWA and LWA respectively. Applying more weight to LWA will lead to a higher estimate, whereas applying more weight to OWA would lead to a lower estimate. This result is slightly counter-intuitive to the goals of the two strategies, as the OWA (focused on down-weighting the lowest judgment) leads to a lower aggregate. This result is due to the fact that the highest judgment is coming from expert 1, who is (by chance in our example) given more weight in the LWA relative to the weight assigned to the highest judgment in the OWA. Indeed, Table 1 shows the SDOWA is in between the LWA and OWA.

This example illustrates one primary objective for combining the LWA and OWA: to shift the relative weights applied to the evidence in such a way as to jointly represent the strategies of the linear ***and*** order weights. The weights are parts of a whole, that represent relative worth. As such, each weight depends on the other weights, creating a dependency between them. The joint consideration of the two sets of weights needs mathematics that accounts for these properties of the weights. In the case of the LWA and OWA, the weights can be represented as compositions, and the mathematics and geometry of compositions can clarify the meaning behind shifting weights along with a deeper understanding of the mechanisms for computing joint weights.

Table 1. Aggregation example

	Expert 1	Expert 2	Expert 3	
Evidence	90.00	50.00	10.00	
Operator weights applied to each expert				Aggregate
LWA	0.60	0.30	0.10	70.00
OWA	0.45	0.50	0.05	66.00
WOWA	0.45	0.42	0.14	62.60
HWA	0.81	0.45	0.02	95.55
SDOWA	0.53	0.40	0.08	68.02
JWA	0.64	0.35	0.01	74.94

3.1 Compositional Geometry

The aim of our paper is to enable the blending of source- and evidence-based strategies towards aggregation. Our proposed approach for this is to blend the

a priori beliefs about the quality of sources and the quality of evidence. This is operationalized by blending the weights assigned to sources (via linear weights) with the weights assigned to the evidence (via order weights) into interpretable joint weights. To achieve this goal, we leverage the fact that non-negative weights that sum to one are compositional, and utilize compositional geometry to combine them.

A composition is formally defined as a collection of components that are constrained to sum to a constant for individual cases (See [1,15,18] for tutorials). Proportions and probabilities are examples of compositions which are constrained to sum to 1. For example, [15] describe how the proportion of the day spent working, relaxing, and exercising (assuming no other activities) could be defined as a composition. If more time is spent working on a given day, then there must be less time spent on relaxing and/or exercising. Therefore, components of a composition cannot be analyzed separately, because the value of one component affects the value of the others.

Compositions have their own geometry defined by a mathematical structure called a simplex. Within the simplex, a composition is defined by each component part. For example, a composition with 3 parts is defined by the set $\{a, b, c\}$ such that

$$a \geq 0; b \geq 0; c \geq 0; \text{and} a + b + c = d.$$

Constraining d to equal 1 generates a standardized simplex that elegantly reflects compositions made up of proportions and probabilities.

Because each component is dependent on the value of the remaining components, the operation of addition is slightly different in compositional geometry. Compositional addition is known as *perturbation*, and defines how the relative values of each component will change when added to another component. Given composition $\mathbf{x} = \{x_1, x_2, x_3\}$ and $\mathbf{y} = \{y_1, y_2, y_3\}$, 'addition' in compositional geometry is defined using the symbol $\oplus$ as,

$$\mathbf{x} \oplus \mathbf{y} = \{\frac{x_1 * y_1}{\sum_{k=1}^{K} x_k * y_k}, \ldots, \frac{x_K * y_K}{\sum_{k=1}^{K} x_k * y_k}\}. \tag{9}$$

In compositional algebra, the additive identity is the uniform composition where $x_k = 1/K$ for all k, where K is the number of parts, for example the number of sources in an aggregation context.

Compositions and perturbations have the following properties,
(1) **scale invariance**: results do not depend on the constant sum.
(2) **permutation invariance**: results do not depend on the order of the parts.
(3) **subcompositional coherence**: results of subcompositions (any subset of component parts) do not differ from the their results when considering the full composition.

These properties make compositional geometry a reasonable mathematical representation for the linear and order based weights. The weights assigned to sources and orders are relative, meaning that the individual weights represent a part of a whole. We can use Eq. (9) as an operator to blend ordered linear

weights with the order weights and the resulting weights are constrained to be non-negative and sum to one.

We therefore propose *joint* weights that are a *perturbation* of the linear and order weights. In general, the perturbation of two quantities with different units produces a result with a joint unit. For example, in [18] (p. 18) the authors outline how a composition of percentages of grams of nutrients for a food item can be transformed to units of energy by adding this composition to a composition that represents the relative kJ/g of each nutrient. Therefore, our joint weights are in the joint unit of source quality and evidence order.

The perturbation of two sets of weights has appeared implicitly in the aggregation literature previously, but its connection to compositional geometry was never identified. It was first introduced by [7] as a way of combining optimism with risk in decision making. It was also considered briefly by [9] as a method to combine order and linear weights, but was ultimately rejected on the grounds that it did not satisfy properties 7 and 8 listed above. Because the connection to compositional geometry was never made, in part because weights have not been considered as individual parts of one whole, the desirable behavior for perturbing and thus systematically and strategically integrating the linear and order weights–as described above–were unknown.

4 The Joint Weighted Average

In this section we introduce the Joint Weighted Average (JWA) as a new aggregation operator that allows for the joint representation of independently defined linear and order weights in computed averages. We therefore propose a new type of blend of LWA and OWA with mathematical properties that differ from the prior literature attempting to combine these two weighting approaches.

Given n sources, let $\mathbf{w} = \{w_1, \ldots, w_n\}$ be a set of convex source weights and let $\mathbf{v} = (v_1, \ldots, v_n)$ be a vector of convex order weights. Both $\mathbf{v}$ and $\mathbf{w}$ are compositions because their components are non-negative and sum to 1. Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be the evidence generated by the n sources. We reorder the evidence and the linear weights by the same permutation function $\pi()$ into respective vectors $\mathbf{x}_\pi$, and $\mathbf{w}_\pi$, such that $x_{\pi(1)} \geq x_{\pi(2)} \geq \ldots, \geq x_{\pi(n)}$. Finally, the ordered linear weights are perturbed by the order weights (using the compositional $\oplus$ operator), and these weights are applied to the ordered evidence to generate a weighted average given by

$$JWA(\boldsymbol{x}) = (\mathbf{w}_\pi \oplus \mathbf{v})\boldsymbol{x}_\pi^T. \tag{10}$$

The operations inside the parentheses represent compositional operations. We then take the inner product of the resulting composition with the data in $\mathbf{x}_\pi$ to calculate the weighted average.

Returning to the example in Table 1, treating the LWA and OWA's weights as compositions can resolve all the problems identified in the other operators. Given that the lowest judgment (which has the lowest priority in the OWA) is generated by expert 3 (who has the lowest priority in the LWA), a joint focus on

both strategies shifts weight away from expert 3's judgment. This logic results in the JWA weights displayed in Table 1, with an even larger aggregated estimate than would result from the LWA, OWA, and any weighted combination of them. The only other aggregator that displays this behavior is the HWA, but this is due to the fact that the ordering of the $\{x_1, x_2, x_3\}$ is the same as the order of $\{w_i x_1, w_2 x_2, w_3 x_3\}$, which is not always the case.

The mathematical properties of the JWA operator can be derived from two basic facts: (1) the vector inside the parentheses is non-negative and sums to one, making the JWA have the same properties as a weighted average and (2) that these weights are derived using well-defined compositional geometry. First, because this operator can be represented as a weighted combination of the data $\mathbf{x}_\pi$, it automatically satisfies properties 1–3. Second, within compositional geometry, the additive identity is the composition with uniform weight on each component. Therefore, if $\boldsymbol{w}_\pi = 1/n$ for all $i \in \{1, \ldots, n\}$ then $\boldsymbol{w}_\pi \oplus \mathbf{v} = \boldsymbol{v}$ and $JWA(\boldsymbol{x}) = OWA(\boldsymbol{x})$. If $\boldsymbol{v} = 1/n$ for all $i \in \{1, \ldots, n\}$ then $\boldsymbol{w}_\pi \oplus \boldsymbol{v} = \boldsymbol{w}_\pi$ and $JWA(\boldsymbol{x}) = LWA(\boldsymbol{x})$. The JWA therefore also satisfies property 6.

In [9], the authors observed violations of monotonicity (property 5) from a similar perturbation of LWA and OWA. While this appears to be problematic behavior, compositional geometry reveals that this is only due to instances where the data ordering changes. In fact, JWA satisfies monotonicity if and only if $\boldsymbol{y}$ has the same ordering as $\boldsymbol{x}$, which is a more reasonable restriction. When the ordering of evidence changes, the interaction between the LWA and OWA weights should change.

Similar to the WOWA and HWA, the JWA does not satisfy properties 7 and 8. For internal boundedness (property 7), the results in Table 1 provide a demonstration that the JWA is not bounded by the OWA and LWA, and can in fact be less than the $min(OWA, LWA)$ or greater than the $max(OWA, LWA)$. For transitivity of equality (property 8), the result of $OWA(\boldsymbol{x}) = LWA(\boldsymbol{x}) = k$ can be achieved for all $\boldsymbol{x} \in R^n$ by having $\mathbf{w}_\pi = \mathbf{v}$, meaning the linear and order weight assigned to each element of $\mathbf{x}_\pi$ is the same for all $i \in \{1, \ldots, n\}$. In this case, for $JWA(\boldsymbol{x}) = k$ we require $\mathbf{w}_\pi = \mathbf{v} = \mathbf{w}_\pi \oplus \mathbf{v}$. However, this condition is only met when $\mathbf{w}_\pi = \mathbf{v} = (1/n, \ldots, 1/n)$. In all other cases, $OWA(\boldsymbol{x}) = LWA(\boldsymbol{x}) \neq JWA(\boldsymbol{x})$.

While these behaviors violate previously desired properties, they are still mathematically sound and interpretable within compositional geometry. In fact, we find that internal boundedness (property 7) and transitivity of equality (property 8) preclude an aggregate that allows the LWA and OWA weights to interact and thus preclude the respective strategies to be blended. For example, evidence receiving weight greater (lesser) than $1/n$ from both the linear and order weights will receive even more (less) weight through perturbation, this desirable behavior and internal boundededness cannot both simultaneously hold.

Conceptualizing the JWA weights as a perturbation of two sets of relative weights requires deeper consideration of the role of weights that equal zero or one. Such values should be assigned with extreme caution. Zero weights mean that no weight will ever be assigned to that component. Therefore the LWA

weights should not be assigned a zero, because a zero weight means that source's evidence should never be considered. Attempting to perturb two sets of weights where the result yields all zeros implies a logical inconsistency in the a priori weights, as the two sets of weights together suggest that none of the evidence should be used. While compositions can contain zeros, a weight of one implies all other weights are zero. Any set of weights for the LWA or OWA that contains a weight equal to one precludes the need for the JWA, as weight can no longer be perturbed in this instance.

5 Experiments

We provide a demonstration of the effectiveness of the JWA operator using simulation. Assume we consult k sources over n trials resulting in a $k \times n$ matrix X that represents the evidence from our sources. For each trial, the aggregate of the source evidence provides a prediction $\hat{y}$ of the criterion variable y.

To show the value of the JWA, we simulate a data set where the quality of the sources **and** the ordering of the data inform on the criterion variable we are trying to predict. We simulate evidence from a collection of sources, and we also simulated the criterion that we are trying to predict. This way we can directly control the statistical properties of the sources, and their evidence, for predicting y. We use a simple model, where the data for X and y are simulated from a joint normal distribution. For a normal distribution, we can set the mean, variance, and covariance of each variable. For simplicity, we set all means and variances equal to the constant 10, so that the worth of the sources is defined by the covariance structure. We set the covariance between sources ($\sigma_{x,x'}$) to the constant 2. We want the sources to have some predictive accuracy, so this is modelled by the degree to which the evidence, X, co-varies with the criterion, y. We adopt the term *validity* from social sciences here to describe the correlation between our observable evidence and the criterion.

The LWA and OWA have weights that reflect the worth of the sources and the evidence in our simulation. Specifically, the worth of the sources is defined in our simulation by changing the relative validity of each source (we hold all other source attributes constant). In other words, the LWA weights directly reflect the relative validity of each source.

To simulate a clear OWA strategy, our simulation injects changes to the interpretation of the evidence by randomly adding a positive bias (or *error*) δ to the evidence generated by two of the sources. We randomly add this bias to 50% of the trials. Since this bias is not a property of any specific source, we randomly select which two sources receive this bias independently for each trial. Therefore, our data context suggests that we should not rely heavily on evidence that is too high as chances are good that this evidence is not a reliable indicator of the criterion y. We therefore set the OWA weights to account for this data context by putting zero weight on the highest two data points with the remaining weights set to $1/(k-2)$. We acknowledge that both LWA and

OWA use prior information in this simulation, the challange here is whether aggregators that combine the LWA and OWA operators can take advantage of this. In other words, the aim of the experiment is not to compare the operators in absolute terms, but to illustrate their individual behavior in reacting to this environment.

Each aggregation operator is applied to the source data X to generate an aggregate that represents a prediction $\hat{y}$ of y. We measure performance as the mean squared error (MSE) between $\hat{y}$ and y across the n trials. We vary the validity and bias parameters as shown in Fig. 1 and Table 2. We report the mean MSE across 500 replications for each parameter combination listed in Table 2.

In our results we show the performance of 4 different aggregation operators that combine the LWA and OWA: (a) JWA, (b) WOWA, (c) HWA, and (d) OWAWA (with $\alpha = 0.5$). For reference, we also show the the performance of the LWA and OWA. We simulate X using the different sets of validity values for the sources displayed in Table 2. For each set of validity values, the LWA weights are equal to these validity values divided by their sum (so they sum to one). Therefore, the LWA operator weights sources by their validity, equally weighting them when the validity of each source is the same, and thus increasing weight on the higher validity sources as the data sets go from 1 (top row of Table 2) to 7 (bottom row in the top half of Table 2).

Figure 1 displays plots of the MSE of each aggregator by the validity sets in Table 2. Each panel reflects a different level of bias as defined by the panel titles. From left to right panel, the bias increases from a small value (within 1 standard deviation of the data in X) to larger values that can more easily stand out relative to the unbiased sources. We can see that the LWA aggregation tends to increase performance as the difference between source validity values increase, but only when the bias term δ is not large. When δ is large, LWA aggregation performs worse as the validity sets increase. This is because the LWA operator effectively focuses on fewer *valid* sources when the sources differ more, and these few valid sources can easily become completely unreliable when they are biased by a large δ. The OWA aggregation is completely unaffected by the changes in both the validity sets and the bias size. This reflects how this aggregator is designed, equally weighting all pieces of evidence, except disregarding the largest two.

The behavior of the JWA aggregation capitalizes on merging the LWA and OWA strategies, showing improved performance as the validity sets increase, yet remaining robust to the size of the bias that alters the data context. Indeed, the JWA operator is able to leverage the LWA and OWA weights to produce a robustly good performance across all bias levels and validity sets. The OWAWA aggregator (which simply averages the output of LWA and OWA) outperforms the JWA aggregator in the left panel, performs similarly to JWA in the middle panel, but performs far worse than JWA in the right hand panel. The left hand panel is a very friendly environment for the LWA, thus benefiting the OWAWA. However, the right hand panel is where the LWA's performance deteriorates due to the harsh data context.

The WOWA and HWA aggregators, that allow the weights to interact, appear to perform worst than the OWAWA aggregator in all cases except validity sets 6 and 7 in the small bias panel. The HWA aggregator deteriorates quickly as the validity sets increase, while the WOWA aggregator deteriorates quickly as the magnitude of the bias increases.

Interestingly, the JWA and OWA aggregates do not change substantially across panels because we selected order weights that would be robust to this specific change in data context. While the LWA aggregate performance deteriorates, the JWA aggregate does not because it relies on the interaction between the LWA and OWA strategies to capitalize on information about source validity while still remaining robust to the random injection of bias. Indeed, the JWA operator shows subtly increased performance as bias increases, as this facilitates the identification and exclusion of the compromised sources, while benefiting from accurate weights for the sources remaining.

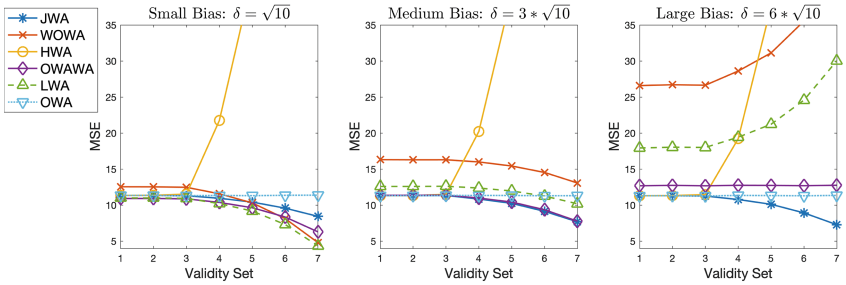


Fig. 1. Performance of operators in aggregating the evidence provided by 10 sources to predict a simulated criterion across different levels of bias and a series of data sets (x-axis), see Table 2. Note: continuous line for visibility only.

Table 2. Variable Simulation Parameters

	Validity $\sigma_{x_i,y}$									
Set	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.91	0.95	0.97	0.99	1.01	1.02	1.03	1.04	1.04	1.05
3	0.68	0.80	0.89	0.96	1.01	1.06	1.10	1.14	1.17	1.20
4	0.03	0.10	0.23	0.42	0.65	0.94	1.27	1.66	2.10	2.60
5	0.00	0.01	0.03	0.10	0.25	0.51	0.95	1.62	2.59	3.95
6	0.00	0.00	0.00	0.00	0.02	0.10	0.34	1.00	2.57	5.96
7	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.23	1.52	8.22

6 Discussion

We have introduced a conceptual framework that recognizes the LWA aggregation and the OWA aggregation as operationalizations of two independent strategies for aggregating data. We consider these strategies to be represented by their respective weights, and build a mathematical framework that appropriately treats the weights of each strategy as parts of a whole in order to blend the strategies. This leads to the use of compositional geometry to blend weights that represent the quality of sources and the quality of the evidence they produce. We have shown that prior work on information aggregation does not jointly account for both the quality of source and the quality of evidence. Through the examples in Table 1 and Fig. 1, we highlight the importance of simultaneously considering both the quality of sources and the quality of information arising from said sources, showing how these can be independently defined and leveraged to improve information aggregation and the handling of inter-source uncertainty. We introduce the Joint Weighted Average (JWA) operator as the first aggregator capable of jointly and systematically focusing on these two features that define the quality of evidence.

The JWA operator jointly focuses on source-and-evidence specific worth by directly combining the linear and ordered weights using compositional geometry [1,14,15]. This approach allows researchers more transparency and a strong mathematical framework to understand the JWA operator's properties and to augment the operator for specific applications. Viewing the weights from the LWA and OWA operators as compositions also naturally leads to better control and understanding of the combined weights allowing more flexibility than previously proposed approaches.

This approach reveals how previous work attempting to combine the LWA and OWA operators have suggested that a combined opperator must adhere to a set of mathematical properties that precludes this joint focus. Specifically, [4,9] propose that the output of aggregators that combine the LWA and OWA operators must be bound between the outputs of the LWA and OWA operators. We show that this property does not allow the strategies of the LWA and OWA operators to be fully merged, and the cost of this property is clearly shown in the right hand panel of Fig. 1. Averaging the outputs of the LWA and OWA operators leads to poor performance if one of the two aggregates faces an environment that degrades their approach, where the JWA operator can withstand such environments by blending their strengths more effectively.

6.1 Conclusions and Future Work

The JWA operator represents a single, simple, interpretable instance of a potentially broad class of weighting strategies that combine individual weighting strategies based on the quality of sources and the quality of evidence. By articulating the compositional nature of this type of combination, and leveraging the well-defined framework of compositional geometry, the JWA operator enables the combination of aggregation strategies in a systematic fashion which has thus

far been out of reach. We note that the JWA operator is also a specific case of an operator that can define the disentanglement of weights for improved estimation and interpretation of aggregation strategies as further discussed in [3].

This provides on the one hand new pathways to understanding how humans combine information and make decisions, and on the other, a mechanism to aggregate information in applications from healthcare to mine-detection, and machine-learning more broadly. Perhaps the most exciting potential lies with the JWA operator providing a bridge linking the processes underpinning human and artificial reasoning, in turn offering more human-centric, understandable, and explainable approaches to AI.

In view of this, we plan to explore how this aggregator might be fit to data to estimate the relative contribution of linear and ordered weights in unknown systems that generate aggregates, such as human judgment–in particular under uncertainty. We believe that this aggregator has the potential to describe psychological processes in a novel way, as well as providing a tool that could automate future aggregations made in this way.

Second, the linear and order weights can be elicited from human users or experts numerically, or indeed as intervals or fuzzy sets–capturing associated uncertainty, to generate an automated information aggregation system with well known mathematical properties. In the above examples, we provided weights for the sources and the evidence that reflected priorities for aggregates. As such the JWA operator can be used to define what information would need to be elicited as well as the implications for what aggregates these two sets of weights would produce, to see if human users agree that the resulting aggregates indeed reflect the priorities given.

Finally, combining linear and order weights plays an important role in XAI by providing parameters that allow people to understand the role of the source vs. the role of evidence in aggregation and act upon it. More broadly, at the level of AI and machine learning, where LWA aggregation approaches are used pervasively, all the way to the heart of neural networks; broadening the perspective and research to explore joint source and evidence weights holds substantial promise.

Acknowledgements. The research underpinning this work was in part supported by the NSF (SES-2343580) and UKRI (ES/Z000084/1) co-funded project *A Novel Theory of Ordered Judgment Processes*.

References

1. Aitchison, J.: The statistical analysis of compositional data. J. R. Stat. Soc., B: Stat. Methodol. **44**(2), 139–177 (1982). http://www.jstor.org/stable/2345821
2. Budescu, D.V., Chen, E.: Identifying expertise to extract the wisdom of crowds. Manag. Sci. **61**(2), 267–280 (2015)
3. Buffum, D., Broomell, S.B., Wagner, C., Anderson, D.T.: The compositional nature of weights. In: Proceedings IPMU (2024)

4. Cardin, M., Giove, S.: SDOWA: A New OWA Operator for Decision Making. Springer Singapore, Singapore (2021). https://doi.org/10.1007/978-981-15-5093-5_28
5. Davis-Stober, C.P., Budescu, D.V., Broomell, S.B., Dana, J.: The composition of optimally wise crowds. Decis. Anal. **12**(3), 130–143 (2015)
6. Davis-Stober, C.P., Budescu, D.V., Dana, J., Broomell, S.B.: When is a crowd wise? Decision **1**(2), 79–101 (2014)
7. Engemann, K.J., Filev, D.P., Yager, R.R.: Modelling decision making using immediate probabilities. Int. J. Gen Syst **24**(3), 281–294 (1996)
8. Huang, S., Broomell, S.B., Golman, R.: A hypothesis test algorithm for determining when weighting individual judgments reliably improves collective accuracy or just adds noise. Decision **11**(1), 7–34 (2024)
9. Llamazares, B.: On generalizations of weighted means and OWA operators. In: Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology, pp. 9–14. Atlantis Press (2011)
10. Merigó, J.M.: On the use of the OWA operator in the weighted average and its application in decision making. In: World Congress on Engineering, vol. 1 (2009)
11. Mesiar, R., Šipeky, L., Gupta, P., LeSheng, J.: Aggregation of OWA operators. IEEE Trans. Fuzzy Syst. **26**(1), 284–291 (2018)
12. Miller, S., Wagner, C., Aickelin, U., Garibaldi, J.M.: Modelling cyber-security experts' decision making processes using aggregation operators. Comput. Secur. **62**, 229–245 (2016). https://doi.org/10.1016/j.cose.2016.08.001
13. Murray, B.J., Islam, M.A., Pinar, A.J., Anderson, D.T., Scott, G.J., Havens, T.C., Keller, J.M.: Explainable AI for the choquet integral. IEEE Trans. Emerg. Top. Comput. Intell. **5**(4), 520–529 (2021). https://doi.org/10.1109/TETCI.2020.3005682
14. Pekaslan, D., Wagner, C.: Compositional linear regression on interval-valued data. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 01–08 (2021). https://doi.org/10.1109/SSCI50451.2021.9660018
15. Smithson, M., Broomell, S.B.: Compositional data analysis tutorial. Psychol. Methods **29**(2), 362–378 (2024). https://doi.org/10.1037/met0000464
16. Torra, V.: Weighted OWA operators for synthesis of information. In: Proceedings of IEEE 5th International Fuzzy Systems, vol. 2, pp. 966–971 (1996). https://doi.org/10.1109/FUZZY.1996.552309
17. Torra, V., Lv, Z.: On the WOWA operator and its interpolation function. Int. J. Intell. Syst. **24**(10), 1039–1056 (2009)
18. Van den Boogaart, K.G., Tolosana-Delgado, R.: Analyzing Compositional Data with R, vol. 122. Springer (2013)
19. Xu, Z., Da, Q.L.: An overview of operators for aggregating information. Int. J. Intell. Syst. **18**(9), 953–969 (2003)
20. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans. Syst. Man. Cybern. **18**(1), 183–190 (1988). https://doi.org/10.1109/21.87068

A Bayesian Modeling Approach to Fuzzy Data Analysis

Antonio Calcagnì[1,2(✉)] and Przemyslaw Grzegorzewski[3,4]

[1] University of Padova, Padua, Italy
[2] GNCS Research Group, National Institute of Advanced Mathematics, Roma, Italy
antonio.calcagni@unipd.it
[3] Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
przemyslaw.grzegorzewski@pw.edu.pl
[4] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Abstract. Statistical data analysis often entails various uncertainties. Fuzzy numbers help address these complexities, enabling generalized statistical methods. We propose enhancing fuzzy estimators by integrating a general epistemic mechanism. Our approach, validated through simulation studies, offers a flexible solution for fuzzy data analysis.

Keywords: fuzzy statistics · gibbs sampler · fuzzy numbers

1 Introduction

Statistical data analysis frequently involves addressing multiple sources of uncertainty simultaneously. This is particularly evident in the analysis of data from social surveys, where both random components (e.g., sample variation) and systematic components (e.g., such as subjective responses) are intertwined [1]. To disentangle these various sources of uncertainty, fuzzy numbers can be employed, and statistical techniques need to be extended to accommodate the fuzzy representation of the data. Within this framework, specialized methods, like fuzzy Expectation-Maximization (fEM), have been devised to ensure estimation and inference. Nonetheless, due to the construction of epistemic fuzzy estimators, they can suffer from excessive variance [3]. In this contribution, we propose a solution which incorporates the general mechanism assumed to underlie the generation of continuous fuzzy numbers into the definition of fuzzy estimators. The idea relies upon the use of a conditional probabilistic framework that connects the parameters of fuzzy numbers (i.e., the systematic component) to the observed statistical model utilized for data analysis. Consequently, estimation and inference are conducted using the Gibbs sampler-based approach, wherein the full conditional distribution is approximated by sampling from a quadratic approximation of the target posterior distribution [4]. This contribution is structured as follows. Section 2 briefly illustrates the problem we are facing alongside

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 59–66, 2024.
https://doi.org/10.1007/978-3-031-65993-5_7

the proposed solution. Section 3 describes the Gibbs sampler derivations whereas Sect. 4 illustrates the results of a short simulation study to check the approximation used into the Gibbs schema. Section 5 concludes this short article by providing a summary of its findings.

2 A Conditional Sampling Schema

2.1 Statement of the Problem

Let $Y_1, \ldots, Y_n$ be n independent continuous random variables and $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)$ a sample of fuzzy observations. The vector $\tilde{\boldsymbol{y}}$ is a blurred version of $\mathbf{y}$ because of epistemic or *post-sampling* uncertainty-based processes. The interest lies in studying $f_{Y_1,\ldots,Y_n}(\mathbf{y}; \boldsymbol{\theta}_{\boldsymbol{y}})$ with the purpose of making inference on $\boldsymbol{\theta}_{\boldsymbol{y}}$ given the fuzzy sample $\tilde{\boldsymbol{y}}$. From the epistemic perspective on fuzzy statistics, fuzzy observations can be conceptualized as stochastic outcomes influenced by varying degrees of non-random and systematic uncertainty, which obscure the actual, unknown realizations $\mathbf{y}$. In this context, each fuzzy observation $\tilde{y}_i$ consists of mode and precision $\{m_i, s_i\}$ of a Beta-type fuzzy number, which is employed as a general template for representing continuous and unimodal fuzzy numbers [1]. Figure 1 shows two instances of the Beta-type fuzzy number.[1]

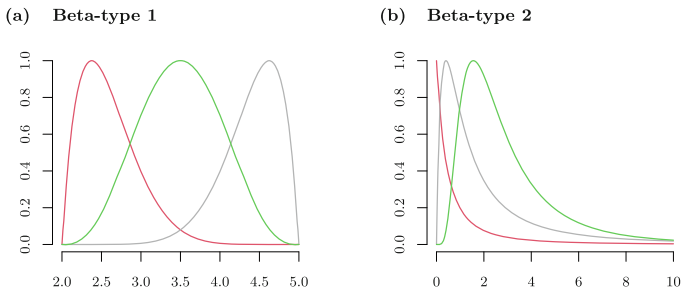


Fig. 1. Some examples of Beta fuzzy sets: (a) bounded case and (b) left-bounded case.

Our objective is to introduce a conditional sampling framework designed to be general enough to handle numerous empirical scenarios involving fuzzy data analysis.

[1] It is important to highlight that Beta-type fuzzy numbers are flexible and parsimonious, requiring only two parameters, while handling with both bounded or semi-bounded variables. Additionally, they serve to generalize triangular fuzzy numbers as well [1].

2.2 Proposed Solution

Our main idea revolves around formalizing the two-stage process that governs the generation of (epistemic) fuzzy numbers. Since these stages interact with each other, the concept involves utilizing a conditional probabilistic framework that connects the parameters of fuzzy numbers (such as modes, spreads, and membership function information if available) to the random outcomes of the model $f_Y(y;\boldsymbol{\theta})$. For the general beta representation of fuzzy numbers, the two-stage sampling framework for fuzzy numbers with modes m_i and precision s_i can be expressed as follows

$$y_i \sim f_Y(y;\boldsymbol{\theta_y}), \tag{1}$$
$$s_i \sim \mathcal{G}a(s;\alpha_s,\beta_s), \tag{2}$$

$$m_i|s_i,y_i \sim \begin{cases} \mathcal{B}e_{4P}(m;s_iy_i,s_i-s_iy_1,lb,ub) & \text{if } y_i \in (lb,ub), \\ \mathcal{B}e_P(m;y_i+y_is_i,s_i+2) & \text{if } y_i \in (0,+\infty). \end{cases} \tag{3}$$

In Eq. (1), $f_Y(Y;\boldsymbol{\theta_y})$ depicts the random variable governing the non-fuzzy sampling process, with its parameters expressed as a function of external covariates $\boldsymbol{\theta_y} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$, as in the case of Generalized Linear Models (GLMs). In Eq. (2), the Gamma random variable $\mathcal{G}a$ with parameters $\alpha_s > 0$ and $\beta_s > 0$ is represented, which models the precision (or spread) of the fuzzy number. In the simplest case, s_i is independent of y_i, although this can be generalized to cases where s_i depends on y_i or external covariates as well. Finally, Eq. (3) represents the random variable governing the mode of the fuzzy number, as a function of the true unobserved outcome y_i and the spread s_i. Depending on the context being modeled, it can be defined as a four-parameter Beta distribution (Be_{4P}) or a Beta-prime distribution (Be_P). Note that, in both cases, with the mode-precision parametrization being adopted, the propagation of fuzziness through Eq. (3) results in $(m_1,\ldots,m_n)$ spreading out near $\mathbb{E}[Y]$. Figure 2 illustrates an example of how fuzziness acts on the true unobserved random realization.[2]

3 Inference on $\boldsymbol{\theta_y}$

Given the conditional sampling structure describing the fuzzification process of the outcomes of $Y_1,\ldots,Y_n$, a natural way to make inference about $\boldsymbol{\theta_y}$ involves a kind of *deblurring* procedure, which uses $\tilde{\boldsymbol{y}}$ instead of the unobserved realizations $\mathbf{y}$. Assuming the observed data is a collection of modes and precisions

[2] Although several choices could have been made for representing the spread (Eq. 2) and mode (Eq. 3) components, the Gamma and Beta distributions exhibit some convenient characteristics in this context. Particularly, Beta distributions, while simplifying the calculus used to obtain the final estimators, also demonstrate great flexibility in modeling unimodal and bounded (or left-bounded) random phenomena. Similarly, the Gamma distribution ensures the positivity of the spread component at a low computational cost.

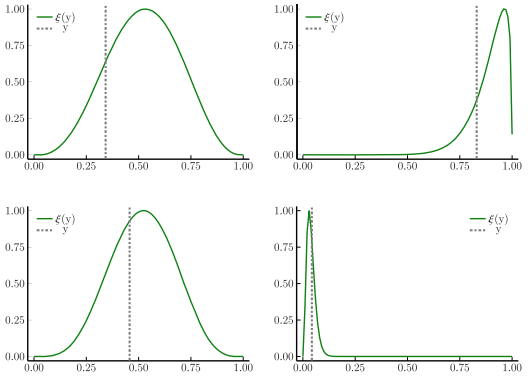


Fig. 2. Examples of the beta-type 1 fuzzy numbers $\xi_{\tilde{y}}$ (green curves) masking the true unobserved realizations y (dashed vertical gray lines).

$\mathbf{D} = (\{m_1, \ldots, m_n\}, \{s_1, \ldots, s_n\})$, the idea is that of using the Gibbs sampler algorithm. It allows for sampling from the posterior distribution $\pi(\boldsymbol{\theta_y}, \mathbf{y}|\mathbf{D})$ by iteratively drawing samples from the conditional posterior distributions ($t > 1$)

$$\mathbf{y}^{(t)} \sim \pi(\mathbf{y}|\mathbf{m}, \mathbf{s}, \boldsymbol{\theta}_y^{(t-1)}), \tag{4}$$

$$\boldsymbol{\theta_y}^{(t)} \sim \pi(\boldsymbol{\theta_y}|\mathbf{m}, \mathbf{s}, \boldsymbol{y}^{(t)}). \tag{5}$$

These are obtained by properly re-arranging the joint distribution as follows

$$\begin{aligned} f(\mathbf{D}, \mathbf{y}; \boldsymbol{\theta_y}) &= f(\mathbf{D}|\mathbf{y}; \boldsymbol{\theta_y}) f(\mathbf{y}; \boldsymbol{\theta_y}) \\ &= f(\mathbf{m}|\mathbf{y}; \boldsymbol{\theta_y}) f(\mathbf{s}|\mathbf{y}; \boldsymbol{\theta_s}) f(\mathbf{y}; \boldsymbol{\theta_y}), \end{aligned}$$

where the prior distribution $f(\boldsymbol{\theta_y})$ has been omitted for the sake of simplicity. It is worth noting that due to the independence between Y_i and S_i, the parameters $\boldsymbol{\theta_s}$ - in the case of Eq. (3), $\boldsymbol{\theta_s} = \{\alpha_s, \beta_s\}$ - can be estimated straightforwardly via maximum likelihood from $\mathbf{s}$. As Eqs. (1)–(3) are quite general and may not lend themselves to analysis within a conjugate Bayesian framework, sampling from the conditional posteriors $\pi(\mathbf{y}|\boldsymbol{\theta_y}, \mathbf{D})$ and $\pi(\boldsymbol{\theta_y}|\mathbf{y}, \mathbf{D})$ could pose challenges. Therefore, we opted to adopt a hybrid solution where the step in Eq. (4) is realized via MCMC (e.g., Block sampling algorithm [5]), the step in Eq. (5) is realized via quadratic posterior approximation [4] instead, as described in Sect. 3.1.

3.1 Approximating $\pi(y|\ldots)$

The quadratic approximation is obtained by matching the parameters of the log posterior density $\pi(y|\ldots)$ with the first K derivatives of a given proposal distribution, which can be chosen according to its feasibility to represent the unnormalized posterior target.

Case 1: $y_i \in (lb, ub)$ In the first case, the proposal posterior density is a four-parameter Beta distribution with mode $\lambda \in (lb, ub)$ and precision $\sigma \in \mathbb{R}^+$. In particular,

$$\ln \pi(y_i|\boldsymbol{\theta_y}, \ldots) \propto \underbrace{-\ln \Gamma(y_i^* s_i) - \ln \Gamma(s_i - s_i y_i^*) + s_i y_i^* \ln\left(\frac{m_i - lb}{ub - m_i}\right)}_{h(y;m,s,lb,ub)} + \ln f_Y(y; \boldsymbol{\theta_y})$$
$$\approx \ln \mathcal{B}e_{4P}(y; \lambda\sigma, \sigma - \sigma\lambda, lb, ub),$$

where the unknown proposal distribution parameters are found by solving the following equations $(K = 2; k = 1, \ldots, K)$

$$\frac{\partial^k}{\partial y^k} \ln \mathcal{B}e_{4P}(y; \lambda\sigma, \sigma - \sigma\lambda, lb, ub) = \frac{\partial^k}{\partial y^k}\Big(h(y; m, s, lb, ub) + \ln f_Y(y; \boldsymbol{\theta_y})\Big),$$

The solutions $\hat{\lambda}$ and $\hat{\sigma}$ are then used into the first Gibbs step

$$y_i^{(t)} \sim Be_{4P}(y; \hat{\lambda}\hat{\sigma}, \hat{\sigma} - \hat{\sigma}\hat{\lambda}, lb, ub).$$

Case 2: $y_i \in (lb, +\infty)$ In the second case, the proposal posterior density is a Beta prime distribution with mode $\lambda \in (lb, +\infty)$ and precision $\sigma \in \mathbb{R}^+$. In particular,

$$\ln \pi(y_i|\boldsymbol{\theta_y}, \ldots) \propto \underbrace{\ln \mathrm{B}(y_i + s_i, s_i + 2)^{-1} + \ln\left(\frac{m_i}{m_i + 1}\right)(y_i + s_i y_i) + \ln m_i + 2\ln(1 + m_i)}_{g(y;m,s)}$$
$$+ \ln f_Y(y; \boldsymbol{\theta_y})$$
$$\approx \ln \mathcal{B}e_P(y; \lambda + \lambda\sigma, \sigma + 2),$$

where the unknown proposal distribution parameters are found by solving the following equations $(K = 2; k = 1, \ldots, K)$

$$\frac{\partial^k}{\partial y^k} \ln Be_P(y; \lambda + \lambda\sigma, \sigma + 2) = \frac{\partial^k}{\partial y^k}\Big(g(y; m, s) + \ln f_Y(y; \boldsymbol{\theta_y})\Big).$$

Similarly to the previous case, the solutions $\hat{\lambda}$ and $\hat{\sigma}$ are used into the first Gibbs step to realize the sampling process

$$y_i^{(t)} \sim Be_P(y; \hat{\lambda} + \hat{\lambda}\hat{\sigma}, \hat{\sigma} + 2).$$

4 Simulation Study

In this section, we briefly present the findings of a simulation study conducted to evaluate the quadratic posterior approximation of $\pi(y|\boldsymbol{\theta_y}, \ldots)$ using $\mathcal{B}e_{4P}$ and $\mathcal{B}e_P$ distributions for both bounded and left-bounded cases.[3] In particular, the derivative-based approximation (DA) has been contrasted against the simple but

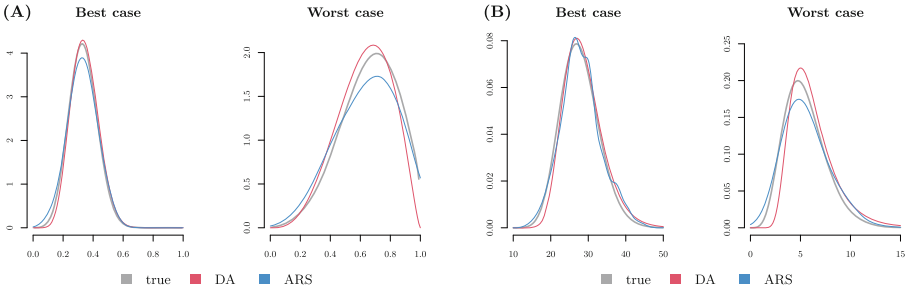


Fig. 3. Simulation study: reconstructed densities (on average) on case 1 (panels A) and case 2 (panels B) in the best and worst result for both DA (red color) and ARS (blue color) algorithms.

still effective Adaptive Rejection Sampling (ARS) algorithm for univariate densities [2]. Two measures have been used to assess the process of the approximation, namely the *computation time* (in log scale) and the *total variation distance*.

Two different designs have been used to evaluate the bounded $y \in (lb, ub)$ and left-bounded $y \in (lb, +\infty)$ cases. In particular, in the first case we have used a Logit-Normal distribution for the non-fuzzy model $f_Y(y; \boldsymbol{\theta_y}) = \mathcal{LG}Norm(y; \mu, \phi)$, with $\mu \in \{-1.85, 0, 1.85\}$ and $\phi \in \{1.0, 3.5\}$. The simplest constraint $lb = 0, ub = 1$ has been considered. Instead, in the second case we have used the Gamma distribution as a model of the non-fuzzy component $f_Y(y; \boldsymbol{\theta_y}) = \mathcal{G}a(y; \alpha, \beta)$, with $\alpha \in \{1.0, 1.5, 3.0\}$ and $\beta \in \{4.9.0, 9.0\}$. In both cases, the independent spread component has been modeled as $s \sim \mathcal{G}a(s; 45.0, 45.0/\mu_s)$, with $\mu_s \in \{5.0, 25.0, 50.0\}$ representing high-to-low fuzziness, whereas $n = 2000$ has been set for each combination of simulation factors.

Figures 4 and 5 show the simulation results across simulation factors for both DA and ARS algorithms. Overall, for the Case 1, the average reconstruction accuracy is 0.97 for DA and 0.95 for ARS, with the average computation log-time being extremely faster for DA (-6.97) then ARS (0.67). Similarly, for the Case 2, the average reconstruction accuracy is 0.93 for DA and 0.85 for ARS, with the average computation log-time being in line with the previous case (DA: -7.27; ARS: 1.15). Finally, Fig. 3 illustrates the reconstructed densities in the best and worst result for both algorithms.

5 Conclusions

In this contribution, we proposed a conditional sampling schema for making inference in statistical analyses based on continuous and unimodal fuzzy data. Fuzzy numbers offer a means to address these complexities, but traditional

[3] The overall results concerning the ability of the proposed schema to efficiently sample from the posterior density $\pi(\boldsymbol{\theta_y}, \mathbf{y}|\mathbf{D})$ are left out for the sake of limited space.

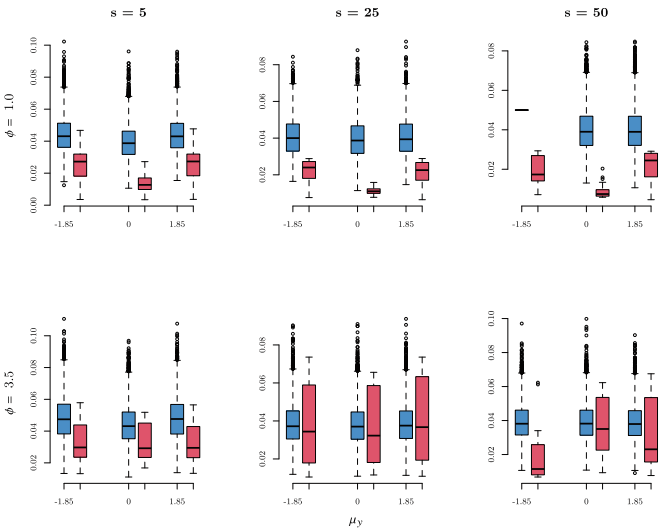


Fig. 4. Simulation study: case 1—total variation distance for each combination of the simulation design for DA (red color) and ARS (blue color) algorithms.

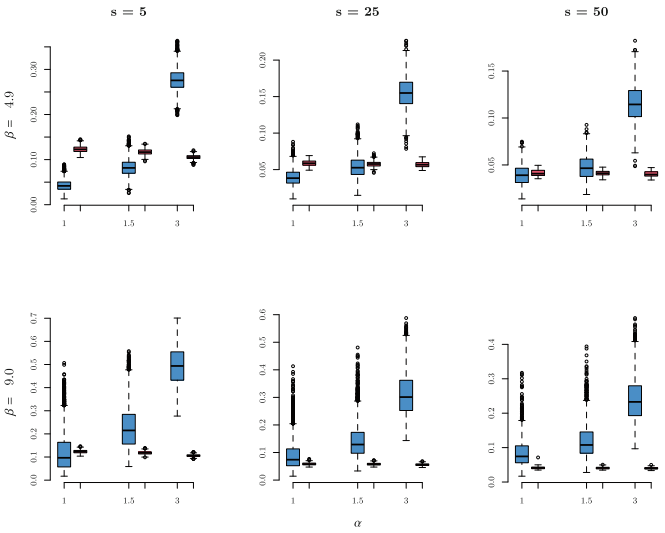


Fig. 5. Simulation study: case 2—total variation distance for each combination of the simulation design for DA (red color) and ARS (blue color) algorithms.

approaches may suffer from high variance. To mitigate this, we propose integrating a general approach which connects fuzzy parameters to observed statistical models, enabling estimation and inference using the Gibbs sampler-based approach. The proposed solution is based on a hybrid Metropolis within Gibbs schema, where a quadratic-based approximation is also used to mitigate the computation burden of the algorithm. A simulation study has been used to assess the quality of the approximation step in two cases of parameters estimation, one involving the Log-Normal distribution and the other involving the Gamma distributions. The results highlights the effectiveness of the proposed solution if compared against the simple but still efficient Adaptive Rejection Sampling algorithm.

Acknowledgements. Antonio Calcagnì acknowledges the financial support provided through the 2022 Italian MUR grant 2022APAFFN "The attentional curve of forgetting visual information in younger and older adults: experimental and computational insights", which supported this research.

References

1. Calcagnì, A., Cao, N., Rubaltelli, E., Lombardi, L.: A psychometric modeling approach to fuzzy rating data. Fuzzy Sets Syst. **44**, 76–99 (2022)
2. Gilks, W.R., Wild, P.: Adaptive rejection sampling for gibbs sampling. J. R. Stat. Soc., C: Appl. Stat. **41**(2), 337–348 (1992)
3. Grzegorzewski, P., Goławska, J.: In search of a precise estimator based on imprecise data. In: 19th World Congress of the International Fuzzy Systems Association (IFSA), 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and 11th International Summer School on Aggregation Operators (AGOP), pp. 530–537. Atlantis Press (2021)
4. Miller, J.W.: Fast and accurate approximation of the full conditional for gamma shape parameters. J. Comput. Graph. Stat. **28**(2), 476–480 (2019)
5. Zhou, H., Huang, X.: Bayesian beta regression for bounded responses with unknown supports. Comput. Stat. Data Anal. **167**, 107345 (2022)

Compound Conditionals and Fuzzy Sets

Lydia Castronovo[1(✉)] and Giuseppe Sanfilippo[2]

[1] Department of Mathematics and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, Messina, Italy
lydia.castronovo@studenti.unime.it
[2] Department of Mathematics and Computer Science, University of Palermo, Palermo, Italy
giuseppe.sanfilippo@unipa.it

Abstract. Following the approach proposed by Coletti and Scozzafava, we propose a slightly different interpretation of fuzzy set theory in terms of conditional events and coherent conditional probabilities, examining it within the framework of the betting scheme. We propose a revised definition of fuzzy sets and introduce operations such as complement and intersection. To achieve this, we employ logical operations among conditional events within the framework of conditional random quantities. Specifically, we interpret the membership function of a fuzzy set as a coherent conditional prevision assessment on suitable compound conditionals. Furthermore, we provide examples and examine the class of Frank t-norms to describe coherent conditional previsions.

Keywords: Conditional events · Conditional random quantities · Conditional prevision · Membership function · Frank t-norms

1 Introduction

In 1965, L. Zadeh [18] proposed the concept of fuzzy set, that is a set characterized by a membership function which assigns to each object a grade of membership. Since that first seminal work, the relationship between probability theory and fuzzy set theory has been widely discussed. Both theories seem to be similar in the sense that both deal with some type of uncertainty and the range of their respective functions is the interval [0, 1]. Zadeh was critical on the difference between probability and fuzzy logic and he described the probability approach as not flexible enough to deal with problems that are usually treated with fuzzy theory. Nevertheless, as pointed out by Zimmerman [19], the comparison between fuzzy set theory and probability theory is difficult primarily for two reasons: first, it strongly depends on which level (mathematically, semantically, linguistically, etc.) such a comparison is made; second both fuzzy theory and probability theory are subject to various interpretations. Indeed, the main criticisms to the probabilistic approach to fuzzy theory took into consideration a "standard" approach to probability. Coletti and Scozzafava [3,4,6], proposed a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 67–75, 2024.
https://doi.org/10.1007/978-3-031-65993-5_8

different approach to fuzzy set theory based on coherent conditional probability. Agreeing that a classical probability approach is not adequate to handle fuzzy logic structures, they considered de Finetti's conditional probability, where conditional events are seen as tri-valued objects that can take values *True*, *False* and *Void*. Through the numerical interpretation of conditional events, probability can model "fuzziness" by expressing partial knowledge in approximate reasoning. In [4], given a property φ of a random quantity X with range C_X, for each $X \in C_X$, the authors focused on the conditional events $E_\varphi|A_x$="You claim (that X has) the property φ, knowing that $(X = x)$". They defined a fuzzy subset of C_X as the set $E^*_\varphi = \{E_\varphi, \mu_\varphi\}$ where $\mu_\varphi(x) = P(E_\varphi|A_x)$ is a coherent conditional probability which measures of how much You, given the event $A_x = (X = x)$, are willing to claim the property φ. The function μ_φ plays the role of the membership function of the fuzzy set E^*_φ. In the present work, we give a similar definition of fuzzy set using coherent conditional probability in the betting scheme framework. Moreover, by exploiting the notion of compound conditional given in the framework of conditional random quantities (see, e.g., [12]), we propose some "set operations" on fuzzy sets. In particular, we interpret the membership functions as previsions of compound conditionals. So the idea is to extend the concept of fuzzy set proposed by Coletti and Scozzafava to the set of pairs (A, μ), where A is a compound conditional (i.e., a conditional random quantity with value in [0,1]) and the membership μ is a coherent conditional prevision on A. The paper is organised as follow. In Sect. 2 we recall some preliminary notions and results. In Sect. 3 we slightly modify the definition of fuzzy set given in [4,6]. Then, we introduce suitable set operations for fuzzy sets. We also analyze the choice of a t-norm as membership of the intersection in the context of coherent conditional probabilities. Finally, in Sect. 4 we give some conclusion and some ideas for future works.

2 Some Preliminary Notions and Results

We use the same symbol to refer to an event A, a two-valued logical entity, and its indicator, which can take value 1, or 0, according to whether A is true, or false, respectively. Given two events A and B, we denote their conjunction as $A \wedge B$ (or AB) and their disjunction as $A \vee B$. Given two events A and H, with $H \neq \emptyset$, the conditional event $A|H$ is a three-valued logical entity which is *true* if AH is true, *false* if $\overline{A}H$ is true, and *void* if $\overline{H}$ is true. The negation $\overline{A|H}$ of a $A|H$ is defined as $\overline{A|H} = \overline{A}|H$. Given a coherent conditional probability assessment $P(A|H) = x$, the indicator of $A|H$, denoted by the same symbol, is $A|H = AH + x\overline{H} = AH + x(1 - H)$. Given a random quantity X, an event $H \neq \emptyset$, and a coherent conditional prevision assessment $\mathbb{P}(X|H) = \mu$, we set $X|H = XH + \mu\overline{H}$ (for more details, see, e.g., [2,12]). We recall the notion of conjunction of two conditional events given in the framework of conditional random quantities [12].

Definition 1. *Given two conditional events $A|H$, $B|K$ and a (coherent) probability assessment $P(A|H) = x$, $P(B|K) = y$, the conjunction $(A|H) \wedge (B|K)$ is defined as the following conditional random quantity $(A|H) \wedge (B|K) = (AHBK + x\overline{H}BK + yAH\overline{K})|(H \vee K)$.*

Notice that the conjunction in Definition 1 can be also represented as $(A|H) \wedge (B|K) = AHBK + x\overline{H}BK + yAH\overline{K} + z\overline{H}\,\overline{K}$, $z = \mathbb{P}[(AHBK + x\overline{H}BK + yAH\overline{K})|(H \vee K)]$.

Theorem 1. *[12, Thm. 7] Given any coherent assessment (x, y) on $\{A|H, B|K\}$, with A, H, B, K logically independent, and with $H \neq \emptyset$, $K \neq \emptyset$, the extension $z = \mathbb{P}[(A|H) \wedge (B|K)]$ is coherent if and only if the Fréchet-Hoeffding bounds are satisfied, that is $z \in [z', z'']$, where*

$$z' = \max\{x + y - 1, 0\}, \quad z'' = \min\{x, y\}. \tag{1}$$

Remark 1. In case of some logical dependencies, for the interval $[z', z'']$ of coherent extensions z it holds that $[z', z''] \subseteq [\max\{x + y - 1, 0\}, \min\{x, y\}]$. In particular, if $HK = \emptyset$, it holds that $(A|H) \wedge (B|K) = (A|H) \cdot (B|K)$ and hence $z = z' = z'' = xy$. See [14] for other cases of logical dependencies.

3 Fuzzy Sets, Complement, and Intersection

In this section starting from the probabilistic interpretation of the membership function of a fuzzy set as a suitable conditional probability given in [3,4], we propose a slightly different definition of a fuzzy subset. Let X be a random quantity with range of values C_X, for each $x \in C_X$ we denote by A_x the event $(X = x)$. Obviously, the family $\{A_x\}_{x \in C_X}$ is a partition (possibly not finite) of the sure event Ω. In this paper we only consider finite sets C_X. Let φ be any property related to the random quantity X. Let us consider an experiment of selecting an agent randomly from a given population. The selected agent, for a given $x \in C_X$, must state whether the property φ holds for X or not. Let E_φ be the following event: E_φ = "They claim that X has the property φ". Then, we consider the conditional event $E_\varphi|A_x$ = "They claim that X has the property φ, knowing that $X = x$" and a conditional probability $P(E_\varphi|A_x)$, seen as a real function $\mu_\varphi(x) = P(E_\varphi|A_x)$ defined on C_X.

Remark 2. We recall that every function $\mu_\varphi : C_X \to [0, 1]$, such that $\mu_\varphi(x) = 0$ if $A_x E_\varphi = \emptyset$, and $\mu_\varphi(x) = 1$ if $A_x \subseteq E_\varphi$, is a coherent conditional probability [4, Theorem 2]. In particular, for every $x \in C_X$, if E_φ and A_x are logically independent, as $A_x E_\varphi \neq \emptyset$ and $A_x \not\subseteq E_\varphi$, it follows that every $\mu_\varphi(x) \in [0, 1]$, $x \in C_X$, is a coherent conditional probability (see also [11, Prop. 12]).

To look at $P(E_\varphi|A_x) = \mu_\varphi(x)$ in the betting framework, we consider a conditional bet that You place on the assertion of the agent randomly chosen from the given population. More precisely, by choosing $P(E_\varphi|A_x) = \mu_\varphi(x)$, for every $s \in \mathbb{R}$, You are willing to pay the amount $s\mu_\varphi(x)$ to receive the random amount

$s \cdot (E_\varphi|A_x) = s \cdot (E_\varphi A_x + \mu_\varphi(x)\overline{A}_x)$. In other words, for every $s \in \mathbb{R}$, You pay the amount $s\mu_\varphi(x)$ in order to receive the amount: s, if They claim that X has property φ and $X = x$; 0, if They claim that X has property $\overline{\varphi}$ and $X = x$; $s\mu_\varphi(x)$, that is the paid amount, if $X \neq x$. Similarly to [4, Definition 2] the following probabilistic interpretation of a fuzzy set can be given.

Definition 2 (Fuzzy set). *Let X be a random quantity with set of possible values C_X and a property φ on X. Given a coherent conditional probability P on the family $\mathcal{F}_\varphi = \{E_\varphi|A_x, x \in C_X\}$, we define a* fuzzy set E^*_φ *in $\mathcal{F}_\varphi$, with membership μ_φ, the set*

$$E^*_\varphi = \{(E_\varphi|A_x, \mu_\varphi(x)) : x \in C_X\}, \text{ where } \mu_\varphi(x) = P(E_\varphi|A_x). \tag{2}$$

In this setting, for every φ, the fuzzy *emptyset*, $\emptyset^*$, is the fuzzy set where the membership function is $\mu_\emptyset(x) = 0$, for every $x \in C_X$. Analogously, the fuzzy sample set Ω^* is the fuzzy set where the membership function is $\mu_\Omega(x) = 1$ for every $x \in C_X$. Of course, the fuzzy set $\emptyset^*$ (resp., Ω^*) can only be defined if the assessment $\mu_\varphi(x) = 0$ (resp., $\mu_\Omega(x) = 1$), $x \in C_X$, is coherent. Based on Remark 2, under the assumption that, for every $x \in C_X$, the events E_φ and A_x are logically independent, it follows that the fuzzy sets $\emptyset^*$ and Ω^* are properly defined. In our framework, the negation of the event E_φ, denoted by $\overline{E}_\varphi$, is the event "They claim that it is not true that X has the property φ". The sentence "it is not true that X has the property φ" will be simply denoted as "X has the property $\overline{\varphi}$". Then, $\overline{E}_\varphi = E_{\overline{\varphi}}$ = "They claim that X has the property $\overline{\varphi}$". Thus, the negation $\overline{E}_\varphi|A_x$ of the conditional event $E_\varphi|A_x$ coincides with the conditional event $E_{\overline{\varphi}}|A_x$="They claim that X has the property $\overline{\varphi}$, knowing that $X = x$".

Definition 3 (Complement of a fuzzy set). *Given a fuzzy set E^*_φ in $\mathcal{F}_X$, the complement of E^*_φ, denoted by $\overline{E}^*_\varphi$, is the following fuzzy set in $\mathcal{F}_{\overline{\varphi}}$: $\overline{E}^*_\varphi = E^*_{\overline{\varphi}} = \{(E_{\overline{\varphi}}|A_x, \mu_{\overline{\varphi}}(x)) : x \in C_X\}$, where $\mu_{\overline{\varphi}}(x) = P(E_{\overline{\varphi}}|A_x)$.*

In Definition 3 coherence requires that $\mu_{\overline{\varphi}}(x) = P(E_{\overline{\varphi}}|A_x) = P(\overline{E}_\varphi|A_x) = 1 - P(E_\varphi|A_x) = 1 - \mu_\varphi(x)$, for every $x \in C_X$. Therefore,

$$\overline{E}^*_\varphi = \{(\overline{E}_\varphi|A_x, 1 - \mu_\varphi(x)) : x \in C_X\}. \tag{3}$$

To illustrate the definition of these fuzzy sets we give the following example.

Example 1. Let X denote the height (in cm) of a certain individual (let us choose Andy), with $X \in C_X = \{140, 141, \ldots, 220\}$. Let φ be the property "tall". We consider the following two events: E_φ = "They claim that Andy is *tall*", A_x= "Andy's height is x cm", $x \in C_X$. Let us consider the fuzzy set E^*_φ on C_X, where, for every $x \in C_X$, $\mu_\varphi(x) = P(E_\varphi|A_x)$ is your degree of belief on the conditional event $E_\varphi|A_x$, that is on the conditional sentence "They claim that Andy is *tall* knowing that his height is x cm". By Definition 3, it holds that the complement $\overline{E}^*_\varphi$ of E^*_φ is the set of the pairs $(E_{\overline{\varphi}}|A_x, \mu_{\overline{\varphi}}(x)) = (\overline{E}_\varphi|A_x, 1 - \mu_\varphi(x))$, for every $x \in C_X$, where $E_{\overline{\varphi}}|A_x$ is the conditional event "They claim that Andy is *not tall*, knowing that his height is x cm".

Based on in Definition 1, given two fuzzy sets E^*_φ and E^*_ψ, we define their intersection $E^*_{\varphi\wedge\psi}$ as follows.

Definition 4 (Intersection). *Let X, Y be two random quantities and φ, ψ two properties on X and Y, respectively. Moreover, let $E^*_\varphi = \{(E_\varphi|A_x, \mu_\varphi(x)) : x \in C_X\}$ and $E^*_\psi = \{(E_\psi|A_y, \mu_\psi(y)) : y \in C_Y\}$ be two fuzzy sets on $\mathcal{F}_\varphi$ and $\mathcal{F}_\psi$, respectively. The intersection $E^*_{\varphi\wedge\psi}$ of E^*_φ and E^*_ψ is the following fuzzy set on $\mathcal{F}_{\varphi\wedge\psi} = \{(E_\varphi|A_x) \wedge (E_\psi|A_y), (x, y) \in C_X \times C_Y\}$:*

$$E^*_{\varphi\wedge\psi} = E^*_\varphi \wedge E^*_\psi = \{((E_\varphi|A_x) \wedge (E_\psi|A_y), \mu_{\varphi\wedge\psi}(x, y)) : (x, y) \in C_X \times C_Y\},$$

where $\mu_{\varphi\wedge\psi}(x, y) = \mathbb{P}[(E_\varphi|A_x) \wedge (E_\psi|A_y)]$.

For each pair $(x, y) \in C_X \times C_Y$, the conjunction $(E_\varphi|A_x) \wedge (E_\psi|A_y)$ in Definition 4 can be interpreted as the compound conditional "They claim that: X has the property φ, knowing that $X = x$, and Y has the property ψ, knowing that $Y = y$".

Remark 3. In [3,4], the authors consider, under suitable assumptions of conditional independence, as conjunction of two conditional events $E_\varphi|A_x$ and $E_\psi|A_y$ the conditional event $(E_\varphi \wedge E_\psi)|(A_x \wedge A_y)$ (Bochvar conjunction), which can be interpreted as the following conditional "They claim that: X has the property φ and Y has the property ψ, knowing that $X = x$ and $Y = y$". Notice that in general the conditional event $(E_\varphi \wedge E_\psi)|(A_x \wedge A_y)$ does not coincide with the conjuntion $(E_\varphi|A_x) \wedge (E_\psi|A_y)$, which is instead a conditional random quantity. Then, the intersection $E^*_\varphi \wedge E^*_\psi$ introduced in Definition 4 differs from the one given in [3, Def. 21, p. 229]. In particular in [3], under the assumption that $P(E_\varphi|(A_x \wedge A_y)) = P(E_\varphi|A_x)$ and $P(E_\psi|(A_x \wedge A_y)) = P(E_\psi|A_y)$, for every $(x, y) \in C_X \times C_Y$, the membership function of the intersection, $\mu'_{\varphi\wedge\psi}$, is interpreted as the probability $P((E_\varphi \wedge E_\psi)|(A_x \wedge A_y))$, for every $(x, y) \in C_X \times C_Y$. Under these hypotheses, it follows that $\mu'_{\varphi\wedge\psi}(x, y)$ satisfies the Fréchet-Hoeffding bounds (1), that is: $\max\{\mu_\varphi(x) + \mu_\psi(y) - 1, 0\} \leq \mu'_{\varphi\wedge\psi}(x, y) \leq \min\{\mu_\varphi(x), \mu_\psi(y)\}$. The assumptions of conditional independence recalled above are essential to have these bounds on the probability of the conjunction. Otherwise, for every $(x, y) \in C_X \times C_Y$, given the assessment $(\mu_\varphi(x), \mu_\psi(y)) \in [0, 1]^2$ on $\{E_\varphi|A_x, E_\psi|A_y\}$, every $\mu'_{\varphi\wedge\psi}(x, y) \in [0, 1]$ is a coherent extension on $(E_\varphi \wedge E_\psi)|(A_x \wedge A_y)$ [16]. We observe that, using the conjunction in Definition 1, that is $(E_\varphi|A_x) \wedge (E_\psi|A_y)$, the Fréchet-Hoeffding bounds (1) are satisfied for every $(\mu_\varphi(x), \mu_\psi(y)) \in [0, 1]^2$. Nevertheless, sometimes the assumption of conditional independence may not be reasonable, as shown in the following example.

Example 2. We set X and E^*_φ as in Example 1. Let Y denote the weight (in kg) of Andy, with $Y \in C_Y = \{40, 41, \ldots, 120\}$. Let ψ be the property "fit". We consider the events E_ψ = "They claim that Andy is *fit*", A'_y= "Andy's weight is y kg", $y \in C_Y$. By Remark 2, every $\mu_\psi(y) \in [0, 1]$ is a coherent assessment for $E_\psi|A'_y$. We observe that it seems natural not to consider E_ψ stochastically independent from A_x conditionally to A'_y, that is, it seems reasonable to assume that $P(E_\psi|A_xA'_y)$

could not coincide with $P(E_\psi|A'_y)$. Indeed, the information on Andy's height could modify the belief on claiming that Andy is fit only knowing his weight. By Definition 4, for instance, the conditional random quantity $(E_\varphi|A_{180}) \wedge (E_\psi|A'_{75})$ can be interpreted as the conjoint sentence "They claim that: Andy is tall, knowing that his height is 180 cm, *and* Andy is fit, knowing that his weight is 75 kg". Suppose to assign the following values to the membership functions: $\mu_\varphi(180) = P(E_\varphi|A_{180}) = 0.85$ and $\mu_\psi(75) = P(E_\psi|A'_{75}) = 0.8$. By Theorem 1, it holds that $\mu_{\varphi\wedge\psi}(180, 75) = \mathbb{P}[(E_\varphi|A_{180}) \wedge (E_\psi|A'_{75})]$ is coherent if and only if $\mu_{\varphi\wedge\psi}(180, 75) \in [\max\{0.85 + 0.8 - 1, 0\}, \min\{0.85, 0.8\}] = [0.65, 0.8]$.

Remark 4. We observe that, by Definition 4, it follows that $E^*_\varphi \wedge E^*_\varphi \neq E^*_\varphi$. Let E^*_φ be a fuzzy set. Then $E^*_\varphi \wedge E^*_\varphi = \{(E_\varphi|A_x) \wedge (E_\varphi|A_{x'}), \mu_{\varphi\wedge\varphi}(x, x')), (x, x') \in C_X \times C_X\}$, where $\mu_{\varphi\wedge\varphi}(x, x') = \mu_\varphi(x)$, if $x = x'$, and $\mu_{\varphi\wedge\varphi}(x, x') = \mu_\varphi(x)\mu_\varphi(x')$, if $x \neq x'$. Indeed, by Definition 1, when $x = x'$, it holds that $(E_\varphi|A_x) \wedge (E_\varphi|A_{x'}) = E_\varphi|A_x$ and hence by coherence $\mu_{\varphi\wedge\varphi}(x, x) = \mu_\varphi(x)$. Otherwise, when we consider pairs $(x, x') \in C_X \times C_X$ where $x \neq x'$, as $A_x A_{x'} = \emptyset$, by Remark 1, it holds that $(E_\varphi|A_x) \wedge (E_\varphi|A_{x'}) = (E_\varphi|A_x) \cdot (E_\varphi|A_{x'})$ and hence, by coherence, $\mu_{\varphi\wedge\varphi}(x, x') = \mu_\varphi(x)\mu_\varphi(x') \leq \min\{\mu_\varphi(x), \mu_\varphi(x')\}$. Therefore, in general, $E^*_\varphi \wedge E^*_\varphi \neq E^*_\varphi$.

It is well-known that, in fuzzy logic, the intersection between a fuzzy set and its complementary set in general does not coincides with the fuzzy empty set. We now observe that this feature still holds when applying Definitions 3 and 4.

Remark 5. Let E^*_φ be a fuzzy set. Then, by Definition 3 and Definition 4, it holds that $E^*_\varphi \wedge \overline{E}^*_\varphi = \{((E_\varphi|A_x) \wedge (E_{\overline{\varphi}}|A_{x'}), \mu_{\varphi\wedge\overline{\varphi}}(x, x')) : (x, x') \in C_X \times C_X\}$, where $\mu_{\varphi\wedge\overline{\varphi}}(x, x') = 0$, if $x = x'$, and $\mu_{\varphi\wedge\overline{\varphi}}(x, x') = \mu_\varphi(x)\mu_{\overline{\varphi}(x')} = \mu_\varphi(x)\mu_\varphi(1 - x')$, if $x \neq x'$. Indeed, we observe that, when $x = x'$, it holds that $(E_\varphi|A_x) \wedge (E_{\overline{\varphi}}|A_x) = (E_\varphi|A_x) \wedge (\overline{E}_\varphi|A_x) = 0$ and hence by coherence $\mu_{\varphi\wedge\overline{\varphi}}(x, x) = 0$. Otherwise, when we consider pairs $(x, x') \in C_X \times C_X$ with $x \neq x'$, by Remark 1 it holds that $(E_\varphi|A_x) \wedge (E_{\overline{\varphi}}|A_{x'}) = (E_\varphi|A_x) \cdot (E_{\overline{\varphi}}|A_{x'})$, which is not constant and can therefore not necessarily be equal to 0. Therefore, as $\mu_{\varphi\wedge\overline{\varphi}}(x, x')$ cannot necessarily be equal to 0, it follows that $E^*_\varphi \wedge E^*_{\overline{\varphi}} \neq \emptyset^*$.

We recall that Definition 1 has been extended to the case of n conditional events in [13]. Definition 4 can be extended to the case of n fuzzy sets as follows.

Definition 5. *Let $X_1, \ldots, X_n$ be n random quantities and n associated properties $\varphi_1, \ldots, \varphi_n$, respectively. Let $E^*_{\varphi_i} = \{(E_{\varphi_i}|A_{x_i}, \mu_{\varphi_i}(x_i)) : x_i \in C_{X_i}\}$ be a fuzzy set on $\mathcal{F}_{\varphi_i}$, $i = 1, \ldots, n$. The fuzzy conjunction $E^*_{\varphi_1} \wedge \ldots \wedge E^*_{\varphi_n}$ is defined as*

$$E^*_{\varphi_1\wedge\ldots\wedge\varphi_n} = E^*_{\varphi_1} \wedge \ldots \wedge E^*_{\varphi_n} =$$
$$= \{((E_{\varphi_1}|A_{x_1}) \wedge \ldots \wedge (E_{\varphi_n}|A_{x_n}), \mu_{\varphi_1\wedge\ldots\wedge\varphi_n}(x_1, \ldots, x_n)) : (x_1, \ldots, x_n) \in C_{X_1} \times \ldots \times C_{X_n}\},$$

where $\mu_{\varphi_1\wedge\ldots\wedge\varphi_n}(x_1, \ldots, x_n) = \mathbb{P}[(E_{\varphi_1}|A_{x_1}) \wedge \ldots \wedge (E_{\varphi_n}|A_{x_n})]$.

In the literature, different operators have been suggested to model the intersection and union of fuzzy sets, such as the max-operator, the min-operator [18], and the bold union [10]. In particular, many authors proposed as aggregation

operators for the intersection and the union of fuzzy sets the so called triangular norms (t-norms) and conorms (t-conorms), respectively [15]. For what concerns the t-norm, with our approach not all of them are acceptable to describe coherent probability for the conjunction of conditional events. A class of t-norms that in cases of logical independence describe coherent prevision assessments for the conjunction of two conditional events is given by the Frank t-norms [14]. They are t-norms T_λ that give a gradual transiction between the Lukasiewicz t-norm T_L and the minimum t-norm T_M [1], that is

$$\max\{x+y-1,0\} \leq T_{\lambda(x,y)} \leq \min\{x,y\}, \lambda \in [0,+\infty]. \tag{4}$$

In particular $T_0(x,y) = T_M(x,y) = \min\{x,y\}$, $T_1(x,y) = T_P(x,y) = xy$, and $T_{+\infty}(x,y) = T_L(x,y) = \max\{x+y-1,0\}$. From (1) and (4) we have that, under logical independence, for every (coherent) pair of conditional probabilities $\mu_\varphi(x) = P(E_\varphi|A_x)$ and $\mu_\psi(y) = P(E_\psi|A_y)$ the assessment $\mu_{\varphi\wedge\psi}(x,y) = T_\lambda(\mu_\varphi(x), \mu_\psi(y))$ is a coherent extension on $(E_\varphi|A_x) \wedge (E_\psi|A_y)$, for every $\lambda \in [0,+\infty]$. Then, Frank t-norms can be used to construct the membership function of the intersection of two fuzzy sets. However, as shown in the example below, the construction of membership functions by means of Frank t-norms for all the intersections which involve three fuzzy sets could lead to case of incoherent conditional prevision [14, Example 7].

Example 3. Given any logically independent events $E_1, E_2, E_3, H_1, H_2, H_3$, we observe that the assessment $(x_1, x_2, x_3) = (0.5, 0.6, 0.7)$ on $\{E_1|H_1, E_2|H_2, E_3|H_3\}$ is coherent. Moreover, as shown in [14, Example 7], the extension $(T_L(x_1,x_2), T_L(x_1,x_3), T_L(x_2,x_3), T_L(x_1,x_2,x_3)) = (0.1, 0.2, 0.3, 0)$ on the family $\{\mathscr{C}_{12}, \mathscr{C}_{13}, \mathscr{C}_{23}, \mathscr{C}_{123}\}$, where $\mathscr{C}_{ij} = (E_i|H_i) \wedge (E_j|H_j)$, and $\mathscr{C}_{123} = (E_1|H_1) \wedge (E_2|H_2) \wedge (E_3|H_3)$, is not coherent. This counterexample shows that, when considering the fuzzy sets $E^*_{\varphi_1}, E^*_{\varphi_2}, E^*_{\varphi_1\wedge\varphi_2}, E^*_{\varphi_1\wedge\varphi_3}$, $E^*_{\varphi_2\wedge\varphi_3}$, and $E^*_{\varphi_1\wedge\varphi_2\wedge\varphi_3}$, the t-norm of Lukasiewics cannot be always chosen as membership function in Definition 5 because it could lead to incoherent assessments.

4 Conclusions

Following the approach introduced by Coletti and Scozzafava in [4], we proposed a framework for fuzzy set theory based on conditional events and coherent conditional probability theory. Given a random quantity X and a property φ related to X, we defined a fuzzy set $E^*_\varphi = \{(E_\varphi|A_x, \mu_\varphi(x)), x \in C_X\}$ where C_X is the range of X and $\mu_\varphi(x) = P(E_\varphi|A_x)$ is a coherent conditional probability on the conditional sentence "They claim that X has the property φ, knowing that $X = x$". Using this definition as a starting point, we introduced suitable operations on fuzzy sets where the membership function is interpreted as the prevision of the corresponding operations on conditional events. In particular, for what concerns the membership of the intersection we analysed that, in the related literature, this is sometimes defined through triangular norms. With our definition of fuzzy set not all the t-norms give a coherent probability assessment on the intersection. A particular role is played by the so called Frank t-norm.

This is just a first step in the formalisation of the membership function as a coherent prevision assessment on compound conditionals. Future work will concern the analysis of other set operations and relations, such as the union, the inclusion of fuzzy sets [5], De Morgan's Laws and the formalization and interpretation of more complex structures such as the "conditional" of two fuzzy sets, $E^*_\varphi | E^*_\psi$ defined through the notion of iterated conditional (see, e.g. [17]). Moreover, based on the fuzzy modal logic for conditional probability proposed in [9], by taking into account the results given in [7,8], the probabilistic interpretation of the membership function can be extended to any Boolean combination of fuzzy sets.

Acknowledgements. We thank the anonymous reviewer for the useful suggestions. Both authors acknowledge support by INdAM-GNAMPA research group and by the MIUR-PRIN project 2022AP3B3B funded by Next Generation EU. G. Sanfilippo is also supported by the FFR2024 project of University of Palermo. Both authors contributed equally to the paper.

References

1. Butnariu, D., Klement, E., Zafrany, S.: On triangular norm-based propositional fuzzy logics. Fuzzy Sets Syst. **69**(3), 241–255 (1995)
2. Castronovo, L., Sanfilippo, G.: A probabilistic analysis of selected notions of iterated conditioning under coherence. Int. J. Approx. Reason. **165**, 109088 (2024)
3. Coletti, G., Scozzafava, R.: Probabilistic Logic in a Coherent Setting. Kluwer, Dordrecht (2002)
4. Coletti, G., Scozzafava, R.: Conditional probability, fuzzy sets, and possibility: a unifying view. Fuzzy Sets Syst. **144**(1), 227–249 (2004)
5. Coletti, G., Scozzafava, R., Vantaggi, B.: Coherent conditional probability, fuzzy inclusion and default rules. In: Soft Computing: State of the Art Theory and Novel Applications. Studies in Fuzziness and Soft Computing, vol. 291, pp. 193–208. Springer, Berlin, Heidelberg (2013)
6. Coletti, G., Scozzafava, R.: Conditional probability and fuzzy information. Comput. Stat. Data Anal. **51**(1), 115–132 (2006)
7. Flaminio, T., Gilio, A., Godo, L., Sanfilippo, G.: Compound conditionals as random quantities and boolean algebras. In: Proceedings of 19th International Conference on Principles of Knowledge Representation and Reasoning, pp. 141–151 (2022)
8. Flaminio, T., Gilio, A., Godo, L., Sanfilippo, G.: On conditional probabilities and their canonical extensions to Boolean algebras of compound conditionals. Int. J. Approx. Reason. **159**, 108943 (2023)
9. Flaminio, T., Godo, L.: A fuzzy probability logic for compound conditionals. In: ESTYLF 20/21, Actas CAEPIA 20/21, pp. 256–261 (2021)
10. Giles, R.: Łukasiewicz logic and fuzzy set theory. Int. J. Man-Machine Studies **8**(3), 313–327 (1976)
11. Gilio, A., Ingrassia, S.: Totally coherent set-valued probability assessments. Kybernetika **34**(1), 3–15 (1998)
12. Gilio, A., Sanfilippo, G.: Conditional random quantities and compounds of conditionals. Stud. Log. **102**(4), 709–729 (2014)
13. Gilio, A., Sanfilippo, G.: Generalized logical operations among conditional events. Appl. Intell. **49**(1), 79–102 (2019)

14. Gilio, A., Sanfilippo, G.: Compound conditionals, Fréchet-Hoeffding bounds, and Frank t-norms. Int. J. Approx. Reason. **136**, 168–200 (2021)
15. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Springer, 1 edn. (2000)
16. Sanfilippo, G.: Lower and upper probability bounds for some conjunctions of two conditional events. In: SUM 2018. Lecture Notes in Computer Science, vol. 11142, pp. 260–275. Springer, Cham (2018)
17. Sanfilippo, G., Pfeifer, N., Over, D., Gilio, A.: Probabilistic inferences from conjoined to iterated conditionals. Int. J. Approx. Reason. 93(C), 103 – 118 (2018)
18. Zadeh, L.: Fuzzy sets. Inform. Contr. **8**(3), 338–353 (1965)
19. Zimmerman, H.: Fuzzy Set Theory and Its Applications. Allied Publishers (1996)

Sharp Polynomial Upper Bound on the Variance

Marco de Angelis(✉)

University of Strathclyde, Glasgow G1 1XJ, UK
marco.de-angelis@strath.ac.uk

Abstract. This short paper presents a new idea leading to a very sharp upper bound on the variance of interval-valued data. The computation of this sharp bound can be carried out in polynomial time in the worst case. For a whole class of interval data this bound is exact as it can be shown that it coincides with the maximum. The algorithm derives from posing the optimisation problem in probabilistic terms, i.e. thinking beyond the deterministic interpretation of an interval. Interval-valued variance can be seen from an imprecise probability's perspective. There are two alternative and non-competing connotations of an interval: a set of real values, and a credal set of all possible probability measures in the given interval. These two connotations do not precipitate any quantitative discrepancies for interval arithmetic. In fact, interval arithmetic provides a means to compute with these imprecise probabilistic objects. This work demonstrates the computational advantage that originates from looking at intervals from an imprecise probabilistic angle, and it may serve as a testimony towards filling the computational void that has for too long discouraged practitioners from computing with interval statistics.

Keywords: Imprecise statistics · interval variance · interval computation

1 Background

Before this work, computing the variance of a data set whose entries are intervals was NP-hard in the general case, e.g. for nested interval data sets [1]. Because the variance is a quadratic form, and because finding the maximum of general quadratic forms has exponential complexity, it is believed that finding the maximum of the variance is exponential too. However if one looks at the problem from a probabilistic standpoint, new exciting polynomial approximations become possible. After all, the variance is not just any quadratic form. The computation of the variance lower and upper bounds for interval-valued data has been extensively studied by Ferson and Kreinovich and documented in [2], who discovered for the lower bound a *quadratic* algorithm in the worst case. In the same study, it is shown that computing the upper bound has exponential complexity in the worst case. Due to such exponential computational barrier, Ferson et al.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 76–84, 2024.
https://doi.org/10.1007/978-3-031-65993-5_9

[2] have resorted to pre-processing strategies for classifying interval-valued data (e.g. non-overlapping, thin, scattered, nested, etc.) to isolate the cases where such an upper bound can be computed in linear or quadratic time [3]. Nested intervals belong to the hardest class: they are known to have a zero lower bound but an upper bound only computable in exponential time. Other cases displaying concurrent overlaps and nesting can also fall in this class. Kreinovich's algorithms implicitly assume that the underlying data are real-valued points, about which there is uncertainty in the form of intervals. Such intervals will be denominated *real-valued* or *point-valued* to emphasise that there is a single point value somewhere in each interval. In this paper, we show that if we go beyond point-valued intervals new more efficient algorithms can be discovered.

2 Problem Statement

The *biased*, or *uncorrected*, or simply *population* variance of a collection of n real-valued numbers x_i is

$$\begin{aligned} V &= \frac{1}{n}\sum (x_i - \mu)^2 \\ &= \frac{1}{n}\sum x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2 . \end{aligned} \tag{1}$$

where $\mu = \sum x_i/n$ is the arithmetic mean. If the data x_i are imprecisely characterised as intervals this formula cannot conveniently be used because of the dependence problem of repeated variables in interval computation.

2.1 Intervals

An interval is denoted $= [\underline{x}, \overline{x}] = \{x \in \mathbb{R} : \underline{x} \le x \le \overline{x}\} \in \mathbb{IR}$, where $\underline{x}$ is the lower bound and $\overline{x}$ is the upper bound and $\mathbb{IR}$ is the space of intervals. The interval arithmetic operations between two intervals $[\underline{x}, \overline{x}]$ and $[\underline{y}, \overline{y}]$ for computing (1) are

$$\begin{aligned} &[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}], \\ &[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}], \\ &[\underline{x}, \overline{x}]^2 = [\max\{0,\ \operatorname{sgn}(\underline{x}\overline{x}) \min\{\underline{x}^2, \overline{x}^2\}\},\ \max\{\underline{x}^2, \overline{x}^2\}]. \end{aligned}$$

While these operations are guaranteed to enclose the true answer in the respective intervals, their use in expressions with repeated variables yields inflated intervals due to the dependence problem [4]. The midpoint and width of an interval are

$$\begin{aligned} m &= \text{mid } [\underline{x}, \overline{x}] = (\underline{x} + \overline{x})/2, \\ w &= \text{wid } [\underline{x}, \overline{x}] = \overline{x} - \underline{x}. \end{aligned}$$

An n-interval $[\mathrm{x}] \in \mathbb{IR}^n$ is a vector whose entries are intervals

$$[\mathrm{x}] = ([\underline{x}_1, \overline{x}_1], ..., [\underline{x}_n, \overline{x}_n]) = ([x_1], ..., [x_n]) .$$

A real-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ whose expression f is evaluated with interval operations satisfies the fundamental theorem of interval analysis. Such an interval extension $[\mathsf{f}] : \mathbb{IR}^n \to \mathbb{IR}^m$ is said to be inclusion monotonic $[\mathsf{f}]([\mathrm{x}]) \subseteq [\mathsf{f}]([\mathrm{y}])$, whenever $[\mathrm{x}] \subseteq [\mathrm{y}]$.

2.2 Quadratic Form

In standard compact matrix form the interval extension of $V = f(\mathrm{x})$ is

$$\mathsf{f}([\mathrm{x}]) = [\mathrm{x}]^T \; Q \; [\mathrm{x}],$$

where $Q = I - E/n$, I is the $n \times n$ identity and E is $n \times n$ matrix of ones. The maximum for V follows from the optimization problem

$$\begin{aligned} \max_{\mathrm{x}} \quad & \mathrm{x}^T \; Q \; \mathrm{x} \\ \text{s.t.} \quad & \underline{\mathrm{x}} \leq \mathrm{x} \leq \overline{\mathrm{x}}. \end{aligned} \tag{2}$$

The quadratic form $\mathrm{x}^T \; Q \; \mathrm{x}$ is convex, thus all its local maxima are located at the corners of $[\mathrm{x}]$. This form makes it easier to see that the maximisation problem (2) is equivalent to an integer programming problem in the space $\{0, 1\}^n$. The global maximum can be found examining all 2^n corners of the search space.

Proposition 1. *The maximum of (2) is attained at one (or possibly more) of the 2^n corners, combination of endpoints, of the search box $[x_1] \times \cdots \times [x_n]$.*

Proof. The proof is given in [2], §Proof of Thm. 3.3, p. 20.

3 Intervals as Imprecise Probabilistic Objects

3.1 Variance of a Dispersive Mixture Distribution

Although an interval is often considered to be a set of real numbers, from the perspective of imprecise probability, an interval can be seen as a credal set, i.e., a set containing all probability distributions with supports strictly inside the interval. For a single interval $[\underline{x}, \overline{x}]$, the best possible estimate of the mean is just $[\underline{x}, \overline{x}]$, while its variance is $[0, (\overline{x} - \underline{x})^2/4]$. In contrast, the variance of a point-valued interval is simply zero. In an imprecise setting, an interval is the set of all possible self bounded probability distributions. The collection of intervals $[\mathrm{x}]$ can be seen as the credal set of all mixtures of bounded probability distributions. The variance of an equal-weight mixture of n distributions with means μ_i and variances σ_i^2 is

$$\begin{aligned} V_{\text{mixture}} &= \frac{1}{n} \sum \left(\sigma_i^2 + \mu_i^2\right) - \left(\frac{1}{n} \sum \mu_i\right)^2 \\ &= \frac{1}{n} \sum \sigma_i^2 + \frac{1}{n} \sum \mu_i^2 - \left(\frac{1}{n} \sum \mu_i\right)^2. \end{aligned} \tag{3}$$

The second line in (3) makes it immediate to see that for point-valued intervals, i.e. when there is certainty about the value within the interval, $\sigma_i = 0$, $\mu_i = x_i$, the expression (3) becomes equivalent to (1). Therefore, a global maximum for the mixture is always an upper bound on the point-valued variance. This can be shown by noticing that the first term of the second line of (3) is always positive, so the variance for the mixture is always larger than the corresponding point-valued variance. Given the integer-programming problem that follows from §2.2 proposition, and as far as the maximum of (3) is concerned, the search is restricted to the space of discrete probability distributions whose mass is placed on the *endpoints* of the mixture.

This makes intuitive sense because the best-possible upper bound for the variance of an interval $(\overline{x} - \underline{x})^2/4$ is given by a 2-discrete distribution with equal masses at the endpoints. These 2-discrete distributions with endpoint mass make up a so called *dispersive mixture* distribution. The mean and variance of each distribution in the dispersive mixture are

$$\begin{aligned} \mu_i &= p_i \ \underline{x}_i + (1 - p_i) \ \overline{x}_i, \\ \sigma_i^2 &= p_i \ (\underline{x}_i - \mu_i)^2 + (1 - p_i) \ (\overline{x}_i - \mu_i)^2, \end{aligned} \tag{4}$$

where $i = 1, ..., n$, $p_i \in [0, 1]$ is the probability mass placed on the left endpoint, and $1 - p_i$ the probability mass on the right endpoint. Combining and rearranging the equations in (4) we have

$$\sigma_i^2 + \mu_i^2 \equiv p_i \ \underline{x}_i^2 + (1 - p_i) \ \overline{x}_i^2, \tag{5}$$

substituting (5) in (3), the variance of the dispersive mixture is

$$V_{\text{dispersive}} = \frac{1}{n} \ \sum \left(p_i \ \underline{x}_i^2 + (1 - p_i) \ \overline{x}_i^2 \right) - \left(\frac{1}{n} \sum p_i \ \underline{x}_i + (1 - p_i) \ \overline{x}_i \right)^2. \tag{6}$$

3.2 Maximum of the Dispersive Mixture Variance

Let $v : [0, 1]^n \to \mathbb{R}$ be the function that maps a vector of masses $\mathrm{p} = (p_1, ..., p_n)$ to the variance (6) of the dispersive mixture $V_{\text{dispersive}} = v(\mathrm{p})$. The maximum of such a function follows from solving the optimisation problem

$$\begin{aligned} \max_{\mathrm{p}} \quad & v(\mathrm{p}) \\ \text{s.t.} \quad & \mathrm{p} \in [0, 1]^n. \end{aligned} \tag{7}$$

This optimisation is different from (2) not just because the domain is different, but also because $v(\mathrm{p})$ is a different quadratic form compared to $f(\mathrm{x})$. Note that the interval extension $v([\mathrm{p}])$ where $[\mathrm{p}] = [0, 1]^n$ is also affected by the dependence problem of interval arithmetic because p is repeated in (6). Let us look at the properties of this quadratic form. The first partial derivatives are

$$\frac{\partial v(\mathrm{p})}{\partial p_k} = \frac{\partial}{\partial p_k}\left(\frac{1}{n}\sum_i \left(p_i\ \underline{x}_i^2 + (1-p_i)\ \overline{x}_i^2\right) - \left(\frac{1}{n}\sum_i p_i\ \underline{x}_i + (1-p_i)\ \overline{x}_i\right)^2\right)$$
$$= \frac{2}{n}\left(\underline{x}_k - \overline{x}_k\right)\left(\frac{\underline{x}_k + \overline{x}_k}{2} + \frac{1}{n}\sum_i p_i\left(\overline{x}_i - \underline{x}_i\right) - \frac{1}{n}\sum_i \overline{x}_i\right). \tag{8}$$

The interval width and midpoint for the k-th component are $w_k = \overline{x}_k - \underline{x}_k$ and $m_k = \left(\underline{x}_k + \overline{x}_k\right)/2$, while the mean upper bound is $\overline{\mu} = \frac{1}{n}\sum \overline{x}_i$. Substituting these quantities in (8), the k-th component of the gradient is

$$\frac{\partial v(\mathrm{p})}{\partial p_k} = -\frac{2}{n} w_k \left(m_k + \frac{1}{n}\sum_i p_i w_i - \overline{\mu}\right). \tag{9}$$

The sign of the first derivative is positive (negative) if the sign of the parenthesis is negative (positive), since $w_k > 0$ for all intervals of the real line

$$\frac{\partial v(\mathrm{p})}{\partial p_k} > 0, \text{ iff}$$
$$m_k < \overline{\mu} - \frac{1}{n}\sum p_i w_i.$$

The second derivative is

$$\frac{\partial^2 v(\mathrm{p})}{\partial p_j p_k} = \frac{\partial}{\partial p_j}\left(-\frac{2}{n} w_k \left(m_k + \frac{1}{n}\sum_i p_i w_i - \overline{\mu}\right)\right)$$
$$= -\frac{2}{n} w_j \left(\frac{1}{n} w_k\right) = -\frac{2\ w_j w_k}{n^2}. \tag{10}$$

Since the second derivative (10) is negative for all values of $\mathrm{p} \in [0,1]^n$, we conclude that $v(\mathrm{p})$ is a concave quadratic form with respect to the mass vector p, thus it admits only a single global maximum in [p]. The concavity and positiveness of the quadratic form ensures that the global maximum is unique.

4 Algorithm

In this section we sketch the algorithm that solves the optimisation problem (7) in polynomial time. Two consecutive tasks are undertaken: 1) a *threshold task*, and 2) a *gradient task*, which is only invoked for a particular outcome of 1).

4.1 Threshold Task

This task labels each problem either to a low or a high complexity class. If the problem has low complexity the algorithm need not make any additional steps

and can terminate returning the global maximum. For low complexity problems, not only the algorithm is fastest, i.e. as fast as computing a dot product of size n, but also returns an exact upper bound on the point values variance. For high complexity problems the gradient task must be invoked. The expression of the first derivative (9) suggests the existence of a "centre of rotation" around which the masses must be spread out for (6) to be maximised. Intuitively, given the discriminating role of the midpoints m_i and the linear dependence on the widths w_i, a strategic centre is the mean midpoint weighted by the widths

$$\tilde{m} = \frac{\sum m_i w_i}{\sum w_i}. \tag{11}$$

The threshold $t(\mathrm{p}) = \overline{\mu} - \frac{1}{n}\sum p_i w_i$ is introduced for the sign of the first derivative

$$\operatorname{sgn} \frac{\partial v(\mathrm{p})}{\partial p_i} = \begin{cases} +1 & m_i < t(\mathrm{p}) \\ -1 & t(\mathrm{p}) < m_i \end{cases}. \tag{12}$$

An initial vector of masses $\mathrm{p}^{(0)}$ can be derived simply by looking at which midpoints lie before or after the centre $\tilde{m}$

$$p_i^{(0)} = \begin{cases} 1 & m_i < \tilde{m} \\ 0 & \tilde{m} < m_i \end{cases}. \tag{13}$$

The algorithm for this threshold task is articulated in the following steps

1. Evaluate $t(\mathrm{p}^{(0)}) = \frac{1}{n}\sum p_i^{(0)} w_i$,
2. Compare $t(\mathrm{p}^{(0)})$ against $\tilde{m}$ and terminate if any of the following is true
   - $t(\mathrm{p}^{(0)}) = \tilde{m}$, or
   - $t(\mathrm{p}^{(0)}) < \tilde{m}$ and $m_i \notin [t(\mathrm{p}^{(0)}),\ \tilde{m}],\ i = \{1, ..., n\}$, or
   - $t(\mathrm{p}^{(0)}) > \tilde{m}$ and $m_i \notin [\tilde{m},\ t(\mathrm{p}^{(0)})],\ i = \{1, ..., n\}$.
3. Otherwise gradient iterations are needed.

If step 2. is successful, $\mathrm{p}^{(0)}$ is the maximal dispersive distribution and $v(\mathrm{p})$ is an exact upper bound for the point values variance. The complexity of this task is dominated by the dot product $t(\mathrm{p})$, which can optimistically be deemed $O(n)$.

4.2 Gradient Task

This task is only invoked if $\mathcal{H} \neq \emptyset$, where $\mathcal{H} = \{h : m_h \in [\underline{d}, \overline{d}] \ \wedge \ m_h = t(p_h)\}$,

$$[\underline{d}, \overline{d}] = \begin{cases} [t(\mathrm{p}^{(0)}),\ \tilde{m}] & t(\mathrm{p}^{(0)}) < \tilde{m} \\ [\tilde{m},\ t(\mathrm{p}^{(0)})] & \tilde{m} < t(\mathrm{p}^{(0)}) \end{cases}. \tag{14}$$

Let $\mathrm{p}^* = (p_1^*, ..., p_n^*)$ be the maximal mass vector, solution of the optimisation (7). Then the following statements are true

- The masses $\mathrm{p}_{\mathcal{H}}^*$ are interior solutions, i.e. $p_h^* \in (0, 1),\ \forall h \in \mathcal{H}$.
- The midpoints $\mathrm{m}_{\mathcal{H}}$ must be all equal.

Collections of symmetric intervals whose midpoints are all equal have the largest cardinality $\#\mathcal{H} = n$ and its solution is $p_h^* = 0.5$ for all $h \in \mathcal{H}$. Collections of intervals whose midpoints are all mutually different have the smallest cardinality $\#\mathcal{H} = 1$. Symmetric or partially symmetric intervals are not uncommon in the applications. The gradient task must only be invoked for $h \in \mathcal{H}$, because only the masses $p_h \in (0, 1)$ need updating. The algorithm articulates as follows

1. Assign $p_i = p_i^{(0)}$.
2. Find the subset of indices $\mathcal{H} \subseteq \{1, ..., n\}$.
3. $\forall i \notin \mathcal{H}$, set $p_i = 1$, for $m_i < m_h$ and $p_i = 0$, for $m_h < m_i$.
4. Seek an interior solution for p_h using gradient ascent.

The algorithm terminates when $|t(\mathrm{p}) - m_h| \leq \varepsilon$ for any one $h \in \mathcal{H}$ or for $|v(p^*) - v(p^{(K)})| < \epsilon$, where ε and ϵ are small positive real numbers, and K is the last iteration. Step 3. implements the gradient step

$$\mathrm{p}^{(k+1)} = \mathrm{p}^{(k)} + \gamma \, \nabla v(\mathrm{p}^{(k)}). \tag{15}$$

The learning rate γ is designed to never make $\mathrm{p}^{(k+1)}$ overshoot outside $[0, 1]$. This is done by normalising the gradient, and considering only the remaining portion of the space, $1 - p_h$ if $\nabla v(p_h) > 0$ or p_h if $\nabla v(p_h) < 0$. This intuition can be implemented in a single line using the indicator function

$$I\mathrm{p} = \begin{cases} 1 & \nabla v(\mathrm{p}) \geq 0 \\ 0 & \nabla v(\mathrm{p}) < 0 \end{cases}. \tag{16}$$

With this engineered learning rate, that is no longer a parameter and need not be an input of the algorithm, the gradient step becomes

$$\mathrm{p}^{(k+1)} = \mathrm{p}^{(k)} + \frac{\nabla v(\mathrm{p}^{(k)})}{||\nabla v(\mathrm{p}^{(k)})||_2} \left((1 - \mathrm{p}^{(k)}) \, I\mathrm{p}^{(k)} \; + \; \mathrm{p}^{(k)} \, (1 - I\mathrm{p}^{(k)}) \right). \tag{17}$$

These iterations are guaranteed to converge because of the concavity of the objective function and the engineered learning rate.

Complexity. Typically the gradient descent for convex functions converges after $O(1/\epsilon)$ iterations in the worst case, where $\epsilon = |v(\mathrm{p}^*) - v(\mathrm{p}^{(K)})|$ Each iteration must only evaluate the gradient $\nabla v(\mathrm{p})$ a dot product. So, a sufficiently small $\epsilon = 10^{-5}$ leads to 10^5 evaluation of dot product. Under stronger regularity conditions, e.g. Lipschitz gradient, strong convexity, convergence can be even faster $O(\log(1/\epsilon))$. Overall the complexity of the algorithm is $O(mn)$, where $m = 1/\epsilon$ or $m = \log(1/\epsilon)$ depends on the quality of the gradient ascent approximation ϵ.

4.3 Examples

For the sake of demonstrating the efficiency of the algorithm, we present three examples: D1) a collection of $n = 9$ intervals with some nesting and overlaps, D2) a collection of $n = 9$ fully nested intervals, D3) a collection of $n = 20$ fully nested intervals with some coinciding midpoints. Table 1 shows that the

algorithm's upper bound is very close to the exact one. Table 2 shows the number of iterations as the size of the data increases, which confirms that the algorithm is $O(mn)$. The slight increase in K with n is due to the fact that more gradient iterations are needed to find $\mathcal{H}$. The example data sets are:

$$
\begin{aligned}
D1 =& \{[3.5, 6.4], [6.9, 8.8], [6.1, 8.4], [2.8, 6.7], [3.5, 9.7], [6.5, 9.9], [0.15, 3.8], \\
& [4.5, 4.9], [7.1, 7.9]\} \\
D2 =& \{[1.0, 9.0], [1.125, 8.25], [1.25, 7.5], [1.375, 6.75], [1.5, 6.0], [0., 10.], \\
& [1.625, 5.25], [1.75, 4.5], [1.875, 3.75], [2.0, 3.0]\} \\
D3 =& \{[-.75, 10.75], [-.5, 10.5], [-.25, 10.25], [-.05, 10.05], [.05, 9.95], [.25, 9.75], \\
& [.5, 9.5], [.75, 9.25], [1., 9.], [1.125, 8.25], [1.25, 7.5], [1.375, 6.75], [1.5, 6], \\
& [1.625, 5.25], [1.75, 4.5], [1.875, 3.75], [2, 3], [2.25, 2.75], [2.495, 2.505]\}
\end{aligned}
$$

Table 1. Interval arithmetic (IA), exact, and proposed sharp upper bound. K is the number of iterations needed by the gradient task to converge to $\epsilon = \varepsilon = 10^{-9}$, and $\#\mathcal{H}$ is the cardinality of the interior solution set.

	IA $O(n)$	Exact $O(2^n)$	Algorithm, $O(mn)$, $\varepsilon = \epsilon = 10^{-9}$			
	$\bar{\mathrm{f}}([\mathrm{x}])$	$f(\mathrm{x}^*)$	$v(\mathrm{p}^{(0)})$	$v(\mathrm{p}^*)$	K	$\#\mathcal{H}$
D_1, $n = 9$	21.744	10.9744	10.9744	10.9744	0	0
D_2, $n = 9$	28.427	9.7257	9.7257	9.7609	0	1
D_3, $n = 20$	57.997	17.5888	16.9293	17.5890	20	10

Table 2. Number of iterations K with increasing n for $\epsilon = \varepsilon = 10^{-9}$

	$n = 9$	$n = 63$	$n = 207$	$n = 369$	$n = 711$	$n = 1053$
D_1	0	0	0	0	0	0
D_2	0	3	1	2	15	12
	$n = 20$	$n = 140$	$n = 460$	$n = 850$	$n = 1580$	$n = 2340$
D_3	20	61	21	33	50	63

5 Conclusions

This short paper has introduced a new idea to unlock very efficient interval computation for the variance. The algorithm presented in this work can compute in polynomial time a sharp approximation of the upper bound on the variance of interval-valued data especially for cases previously only deemed NP-hard. This work concludes an important chapter on computing with interval statistics that has been open now for about two

decades, and it creates exciting computational prospects for imprecise statistics and its applications. May this idea ignite future imprecise probability research and many more of these algorithms appear in the literature to address other challenging descriptive statistics.

Acknowledgements. Thanks to Scott Ferson and the anonymous reviewer.

References

1. Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., Aviles, M.: Computing variance for interval data is NP-hard. In: ACM SIGACT News, pp. 108–118 (2002)
2. Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., Aviles, M.: Exact bounds on sample variance of interval data. In: Extended Abstracts of 2002 SIAM Workshop on Validated Computing, Toronto, Canada, pp. 67–69 (2002)
3. Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., Ginzburg, L.: Experimental uncertainty estimation and statistics for data having interval uncertainty. Sandia National Laboratories, Report SAND2007-0939 (2007)
4. Moore, R.: Methods and applications of interval analysis. Stud. Appl. Numer. Math. SIAM (1979)

Hierarchical Clustering and CoClust Algorithm: A Nested Procedure to Analyse Sustainable Heating Data

F. Marta L. Di Lascio[1(✉)] and Roberta Pappadà[2]

[1] Faculty of Economics and Management, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy
marta.dilascio@unibz.it

[2] Department of Economics, Business, Mathematics and Statistics "B. de Finetti", University of Trieste, Trieste, Italy
rpappada@units.it

Abstract. In this work we combine two different clustering methods to investigate district heating data by taking into account both static and dynamic information concerning the buildings energy profile. The idea is to use the hierarchical clustering algorithm based on the Gower's index to find a first partition of buildings based on their static characteristics, such as age class, energy class, and heating surface, and, next, to investigate the within-cluster multivariate dependence of thermal energy demand among buildings. The two-step procedure we propose aims at assessing the usefulness of static information to support the management of energy demand in the urban area. We show the procedure on data concerning the district heating system of the Italian city Bozen-Bolzano.

Keywords: CoClust · hierarchical clustering · thermal energy demand

1 Introduction

The study of thermal energy demand (TED hereafter) in urban areas is a crucial need to understand how to increase the efficiency of energy production and distribution systems and improve the sustainability of cities in a way that also has an impact on reducing climate change (see [12]). In recent years, a number of methodologies for analysing thermal demand data have been proposed, mainly based on copula theory (see [9,14] and references therein) and, more recently, on clustering methods (see [10] and references therein).

The first attempt was made by [4] who provided the copula-based probability law of TED taking into account its relationship with meteorological variables, such as outdoor temperature and solar radiation, and derived useful implications for the management and production of thermal energy especially when extreme weather conditions occur. The authors have also investigated different approaches based on classical and rotated copula models (see [4]), vine copulas (see [5]), and convex mixtures of copulas (see [6]).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 85–92, 2024.
https://doi.org/10.1007/978-3-031-65993-5_10

Due to the complexity of the phenomenon investigated, [8] have recently exploited copulas within a clustering framework also taking into account static information of buildings in the urban areas, such as energy class, age class, and heating surface. The authors proposed the use of the Ali-Mikhail-Haq copula (see [1]) to define a dissimilarity measure useful to cluster residential users (buildings) based only on their historical heat demand information, e.g. observed time series of TED, showing the ability of the proposed clustering methodology to identify clusters that are homogeneous in terms of static building characteristics. Finally, [7] refined the study of the impact that meteorological variables have on TED by merging the proposals in [4] and [8]. Thus, the authors enabled the characterization of the probability law of TED (under extreme weather conditions) for specific groups of buildings contributing to the study of the impact of meteorological conditions on the energy needs of buildings in the urban area.

In this paper we use a clustering approach that combines two different methods, a distance-based algorithm and one model-based algorithm, aiming at investigating both the dynamic information, i.e. TED time series, and the static one, e.g. buildings characteristics, related to the district heating (DH hereafter) system of the Italian city Bozen-Bolzano. The final aim is to assess whether the static characteristics of buildings can be exploited in combination with the dependence structure in TED data to support demand forecasting and to ease the management of energy production and distribution.

The remainder of the article is organized as follows. We describe the methodology based on nesting two clustering methods in Sect. 2. In Sect. 3 we describe the data analysed and the empirical results obtained on a specific case study. Conclusions are outlined in Sect. 4.

2 Hierarchical Clustering and CoClust: A Nested Approach

Here we describe the methodology applied to the DH system data analysed in Sect. 3. We propose a two-step procedure that uses different clustering methods as well as data in the two steps by combining the hierarchical clustering algorithm and the CoClust algorithm in [3] in a nested way. More precisely, in a first step, the hierarchical clustering algorithm based on the Gower's index and the complete linkage is applied to static features of different types, e.g. the static characteristics of buildings in a urban area, to obtain a partition of units; in the second step, a within-cluster analysis is carried out by applying the CoClust algorithm to serially uncorrelated time series, e.g. detrended TED time series, of all the units belonging to the same cluster identified in the first step.

Such a procedure allows us to first identify a partition of units (e.g. buildings) that appears to be similar in their characteristics and, then, to assess whether the units belonging to the same cluster show a multivariate dependence in terms of a variable of interest (e.g. TED).

Let $\mathbf{X} = (x_{tj})$, $j = 1, \ldots, d$, be a $(T \times d)$-data matrix, whose columns represent d time series observed at time $t = 1, \ldots, T$ (e.g. TED time series). Assume

also additional information on the units, represented by a $(d \times p)$-data matrix $\mathbf{V}$ of p features available for each time series $\mathbf{x}_j$ $(j = 1, \ldots d)$ in $\mathbf{X}$. The variables $V_1, \ldots, V_p$ can be of mixed type and represent some static information associated to each location (e.g. age, heating surface, energy class). For any pair of units, (i, j), $i, j = 1, \ldots, d$, the well-known Gower's similarity coefficient s_{ij} can be computed from the p features $V_1, \ldots, V_p$, since it can accommodate mixed types of data. Such a coefficient is computed as a weighted sum of all the variable-specific similarities, which are standardized to assume values in [0,1]. This coefficient is converted to dissimilarity by $d_{ij} = 1 - s_{ij}$ (see [10] for details). The obtained $(d \times d)$-matrix can be used as input for an agglomerative hierarchical clustering (AHC) algorithm with a suitable linkage method.

In the second phase, the CoClust algorithm is applied to the partition obtained from AHC. The CoClust is a clustering algorithm that allows the identification of dependent clusters according to their multivariate structure without any assumption on the margins. The kind and the strength of the dependence between clusters are modeled through a copula function. The algorithm groups observations by using the maximized log-likelihood function of a copula fit. The parameter vector (dependence parameter plus marginal parameters) is estimated through the semi-parametric version of the two-stage estimation method called inference for the margins as in [11]. Finally, the CoClust selects automatically the number of clusters by using a criterion based on the log-likelihood of a copula fit. In the proposed procedure, the CoClust allows us to investigate the within-cluster homogeneity in terms of multivariate dependence among the units (e.g. buildings) clustered together according to the static information.

The computational steps of the proposed clustering method are described in the Algorithm 1.

3 A Case Study on Thermal Energy Demand

3.1 The Data

The data analysed in this work concern the TED time series and the static characteristics of the residential buildings connected to the district heating (DH hereafter) system of the Italian city Bozen-Bolzano. Specifically, we use high-frequency TED data (in kWh) of 41 residential users (one or more than one building with homogeneous characteristics fed by one or more DH substations) observed during the two intermediate weeks of January 2016, and detected by 200 heat exchanger substations (each endowed with a smart heat meter) distributed over the urban area of the city. The provided time series have been preprocessed to remove outliers due to meter or transmission system failures and then aggregated to obtain hourly observations. As for the static buildings characteristics, we use heating surface (in dam2), age class (re-coded with 7 levels: 1946–1960, 1961–1970, 1971–1980, 1981–1990, 1991–2000, 2000–2005, >2005), and energy class (re-coded with 5 levels: 71–90 kWh, 91–110 kWh, 111–130 kWh, 131–150 kWh, >150 kWh).

Table 1. Nested AHC-CoClust Algorithm.

Algorithm 1

1. Given the $(T \times d)$ temporal matrix $\mathbf{X}$ and the $(d \times p)$ matrix $\mathbf{V}$ of static features (e.g. buildings characteristics),
2. compute the dissimilarity matrix based on Gower's index $\mathbf{\Delta} = (d_{ij})$, for $i, j = 1, \dots d$, with elements $d_{ij} = 1 - \frac{\sum_{k=1}^{p} s_{ijk} w_{ijk}}{\sum_{k=1}^{p} w_{ijk}}$, where $s_{ijk} \in [0, 1]$ is the similarity between the i-th and j-th unit based on the k-th variable, and w_{ijk} is the weight assigned to the comparison on the k-th variable (typically, $w_{ijk} = 1$, for all i, j, k);
3. **AHC**
   Run an agglomerative hierarchical clustering algorithm with complete linkage (or an alternative linkage method) on Δ and select the number of clusters K via a suitable criterion (e.g. Average Silhouette Width). Denote with g_k the k-th cluster $(k = 1, \dots, K)$;
4. extract the residual (uncorrelated) TED time series $\boldsymbol{\epsilon}'_j = (\epsilon_{1j}, \dots, \epsilon_{tj}, \dots, \epsilon_{Tj})$, $j = 1, \dots, d$, after the fit of appropriated ARIMA models;
5. **CoClust**
   For $k = 1, \dots, K$
   5.1 Let $I^k \subset \{1, \dots, d\}$ be the set of indices of units belonging to g_k, $|I^k| = n_k$;
   5.2 Set $\max(\texttt{dimset}) = n_k$;
   5.3 Run `CoClust` algorithm on $\mathbf{U} = \mathbf{Z}'$ obtained as the transposed of $\mathbf{Z} = (\epsilon_{tm})$, $m \in I^k$;
   5.4 Obtain the final clustering for g_k, $(G_1^k, \dots, G_l^k)$, with $2 \leq l \leq n_k$, based on the maximized log-likelihood of the estimated copula model.

3.2 Empirical Results

Hierarchical clustering of the buildings static characteristics based on the Gower's index leaded to a partition of $K = 6$ clusters selected via the average silhouette width (see [13]).

Figure 1 shows that three clusters (cluster 1, 4, and 6) appears perfectly homogeneous in terms of age class and energy class, and are also well-separated with respect to the heating surface. Cluster 3 is highly homogeneous in terms of energy class and heat surface, whereas it is less compact in terms of buildings age. The remaining two clusters (2 and 5) present a larger heterogeneity compared to the other clusters, although all the members in cluster 2 and cluster 5 share the same energy class and age class, respectively.

The 41 TED time series are preliminary detrended through the fit of appropriated ARIMA models (see [2]) and the (uncorrelated) residuals are extracted. Then the CoClust is applied to the residual TED series of the buildings in each cluster obtained via the first step of Algorithm 1. This leads to the identification, in most of the cases, of a strong dependence relationship among the TED of buildings already clustered by the hierarchical algorithm (see Table 2). The

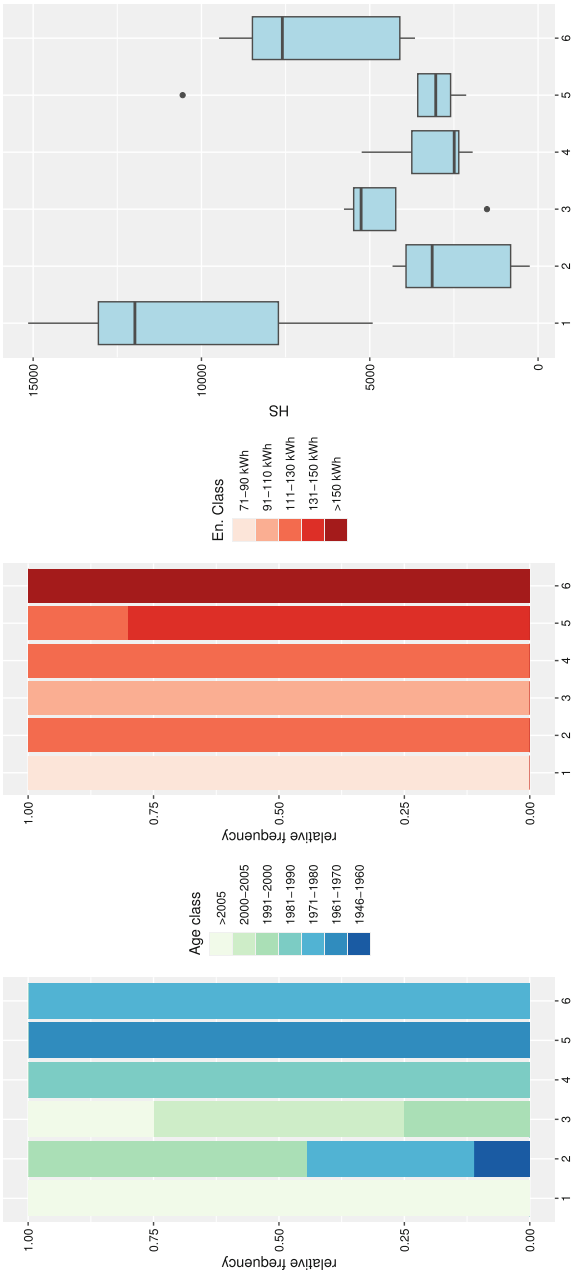


Fig. 1. Age class (left panel), energy class (middle panel), and heating surface (right panel) for each cluster, from 1 to 6, identified by the hierarchical clustering of buildings characteristics on Gower's index and complete linkage

buildings grouped by the CoClust algorithm along with the buildings removed from the cluster membership are reported on the map in Fig. 2.

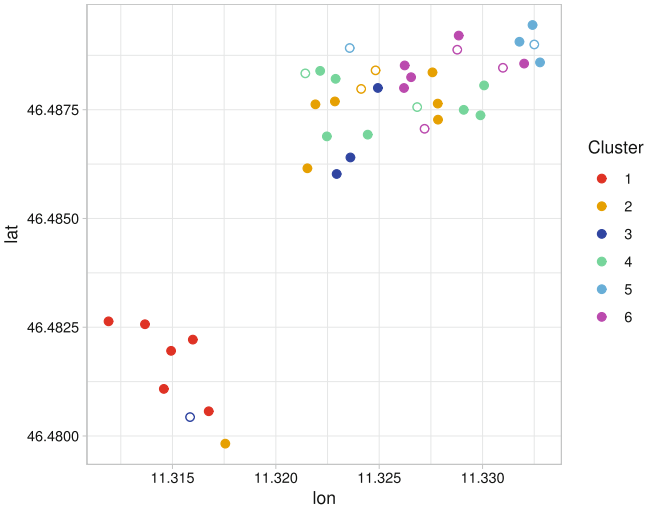


Fig. 2. Residential users partition according to the hierarchical clustering results. The points marked with an empty circle are the buildings dropped out by the CoClust algorithm that grouped together all the remaining buildings in each cluster identified in the HAC step

The results confirm that static characteristics can be considered as an indicator of TED behavior, since the CoClust algorithm is able to characterize the within-cluster dependence structure, with few cases of buildings that are not allocated in the final partition. The identification of buildings that are dropped out by the CoClust algorithm may also provide useful information on TED profile, regardless of the static features, and deserves further investigation.

4 Conclusion

We proposed a nested clustering procedure that exploits the hierarchical clustering method to first identify homogeneous groups of units with respect to some static features; then, a copula-based clustering algorithm, namely the CoClust algorithm, is used to investigate the within-cluster dependence structure on the TED time series of residuals available for each unit. The proposed procedure allows us to (*i*) validate the partition obtained by using only static information, (*ii*) uncover potential heterogeneity among buildings TED profile in the same cluster, and (*iii*) provide insights to define an energy efficiency profile for each cluster in such a way that the buildings static characteristics could be exploited to get instead information on TED.

Table 2. Results of the CoClust algorithm applied to the clusters g_k $(k = 1, \ldots, 6)$ obtained via AHC.

Cluster (AHC)	n. units (AHC)	n. units (Coclust)	Copula model	Copula parameter
g_1	6	6	Frank	1.244
g_2	9	7	Frank	4.148
g_3	4	3	Frank	4.715
g_4	9	7	Frank	4.224
g_5	5	3	Frank	5.176
g_6	8	5	Frank	5.352

Acknowledgements. The authors have carried out this study within the MUR-PRIN 2022 Project "Modelling Non-standard data and Extremes in Multivariate Environmental Time series (MNEMET)" (No. 20223CEZSR) funded by European Union - Next Generation EU. The authors acknowledge Alperia S.p.A. - a company that produces and distributes energy from renewable sources - and the Bozen-Bolzano province for providing the analyzed data on thermal demand.

References

1. Ali, M., Mikhail, N.N., Haq, M.S.: A class of bivariate distributions including the bivariate logistic. J. Multivar. Anal. **8**(3), 405–412 (1978)
2. Box, G., Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1970)
3. Di Lascio, F.M.L., Giannerini, S.: Clustering dependent observations with copula functions. Stat. Pap. **60**(1), 35–51 (2019)
4. Di Lascio, F.M.L., Menapace, A., Righetti, M.: Joint and conditional dependence modelling of peak district heating demand and outdoor temperature: a copula-based approach. Stat. Methods Appt. **29**, 373–395 (2020)
5. Di Lascio, F.M.L., Menapace, A.: Analysis of district heating demand through different copula-based approaches. In: Pollice, A., Salvati, N., Schirripa Spagnolo, F. (eds.) Book of Short Papers—SIS 2020, pp. 292–297. Pearson, Italy (2020)
6. Di Lascio, F.M.L., Menapace, A., Righetti, M.: Analysing the relationship between district heating demand and weather conditions through conditional mixture copula. Environ. Ecol. Stat. **28**(1), 53–72 (2021)
7. Di Lascio, F.M.L., Pappadà, R.: Cluster analysis and conditional copula: a joint approach to analyse energy demand. In: Rampichini, C., La Rocca, M., Coretto, P., Giordano, G., Parrella, M.L. (eds.) Boof ok Abstract and Short Papers—14th Scientific Meeting of the Classification and Data Analysis Group, pp. 427–430. Pearson, Italy (2023)
8. Di Lascio, F.M.L., Menapace, A., Pappadà, R.: A spatially-weighted AMH copula-based dissimilarity measure for clustering variables: an application to urban thermal efficiency. Environmetrics **35**(1), e2828 (2024)
9. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton (2015)

10. Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Cluster analysis, 5th edn. Wiley, Ltd (2011)
11. Genest, C., Ghoudi, K., Rivest, L.P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika **82**, 543–55 (1995)
12. Menapace, A., Santopietro, S., Gargano, R., Righetti, M.: Stochastic generation of district heat load. Energies **14**(17), 5344 (2021)
13. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comp. Appl. Math. **20**, 53–65 (1987)
14. Trivedi, P.K., Zimmer, D.M.: Copula modeling: an introduction for practitioners. Found. Trends Econom. **1**(1), 1–111 (2005)

Fuzzy Clustering Implementations for Big Data in R

Vincenzo Di Perna and Maria Brigida Ferraro(✉)

Department of Statistical Sciences, Sapienza University of Rome,p.le A. Moro 5, 00185 Rome, Italy
diperna.1991664@studenti.uniroma1.it, mariabrigida.ferraro@uniroma1.it

Abstract. The exponential growth in data volume, speed, and variety presents both unprecedented opportunities and challenges across diverse domains. Hence, it is imperative to refine the methodologies and to address the intricacies inherent in the analysis of massive datasets. While implementations of the fuzzy k-means algorithm and its variants are provided by numerous R packages, their computation for extensive datasets demand a considerable amount of time. This inefficiency is not unique to such algorithms, as numerous statistical techniques in R lack the ability to leverage modern resources for computational time reduction. The proposed implementations are designed to enhance the efficiency of the fuzzy k-means type clustering algorithms within the R environment through the integration of parallel computing techniques.

Keywords: R functions · fuzzy clustering · big data · parallel computing

1 Fuzzy k-Means and Its Variants

In a fuzzy framework, the most known clustering algorithm is the fuzzy k-means (fkm) [2], the generalization of the well-known k-means algorithm. To partition n objects contained in the data matrix $\mathbf{X}$ into k clusters, the fkm algorithm aims at solving the following problem:

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{H}} J_{\mathrm{f}k\mathrm{m}} = \sum_{i=1}^{n}\sum_{g=1}^{k} u_{ig}^{m} d^2\left(\mathbf{x}_i, \mathbf{h}_g\right), \\ &\text{s.t.} \quad u_{ig} \in [0,1], \sum_{g=1}^{k} u_{ig} = 1, \end{aligned} \tag{1}$$

where $d\left(\mathbf{x}_i, \mathbf{h}_g\right)$ is the Euclidean distance. In (1) $\mathbf{U}$ is the $(n \times k)$ membership matrix where each entry u_{ig} represents the membership degree of object i to cluster g, taking values in the interval $[0,1]$. The $(k \times p)$ matrix $\mathbf{H}$ contains the prototypes (centroids) $\mathbf{h}_g^{'} = [h_{g1}, h_{g2}, \cdots, h_{gp}]$ $(g = 1, \cdots, k)$ and m is the parameter of fuzziness.

The optimal solution can be found by means of an iterative algorithm. The update equations, obtained by means of the Lagrangian multiplier method, are:

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 93–101, 2024.
https://doi.org/10.1007/978-3-031-65993-5_11

$$\mathbf{h}_g^{'} = \frac{\sum_{i=1}^{n} u_{ig}^m \mathbf{x}_i}{\sum_{i=1}^{n} u_{ig}^m}, \quad g = 1, \cdots, k. \tag{2}$$

and

$$u_{ig} = \frac{1}{\sum_{g'=1}^{k} \left(\frac{d^2(\mathbf{x}_i, \mathbf{h}_g)}{d^2(\mathbf{x}_i, \mathbf{h}_{g'})} \right)^{\frac{1}{m-1}}}, \quad i = 1, \cdots, n, \quad g = 1, \cdots, k. \tag{3}$$

A variant of f*k*m has been introduced in [15]. The proposal consists in avoiding the parameter of fuzziness m and introducing an entropic regularization in the minimization function in order to obtain a fuzzy partition. Since in the f*k*m and the entropic f*k*m the Euclidean distance is used, the obtained partition in both cases contains spherical-shape clusters. In order to overcome this drawback the Euclidean distance may be replaced by a different distance measure. Gustafson and Kessel [10] propose to replace in (1) the Euclidean distance with a cluster-specific Mahalanobis distance. An entropic variant of the Gustafson and Kessel proposal has been introduced in [8]. For a review on fuzzy clustering algorithms refer to [5].

In recent years, datasets have become so large and complex that it is difficult to work with the available database management tools. Fuzzy clustering proposals have been introduced to handle large datasets. Some of them use sampling techniques to decrease the dataset size [3,11], others are based on implementations with Apache Hadoop, Apache Spark or MapReduce [9,12,13,17].

In this paper we focus on `R`. To the best of our knowledge, the f*k*m method is implemented in the following `R` packages: `fclust` [6,7], `e1071` [16], `ppclust` [4] and `advclust` [1]. In particular, the `fclust` package contains several algorithms for fuzzy clustering, cluster validity indices and plots for cluster validity and visualizing fuzzy clustering results. Unfortunately, the above-mentioned `R` packages are very time-consuming when dealing with large datasets. One of the main problems in `R` is the memory limitation. Therefore, it is not possible to store larger data in memory. It is unrealistic to handle data larger than the available RAM because it slows down performance dramatically. Since clustering is essential for the 'Big Data' problem, we introduce new implementations of fuzzy clustering methods for large datasets that can overcome the memory limitation of `R`.

2 Accelerating Data Analysis

The `R` software uses data frames, vectors, arrays and lists as fundamental structures for organising data. These structures are directly stored in RAM,[1] enabling

[1] Random Access Memory, or RAM, is a type of electronic computer memory capable of accessing data in uniform time intervals, regardless of the data's physical location within the memory.

efficient data retrieval and manipulation. When any of these structures is modified, the R programming environment generates a copy of it in memory. This operational behaviour can lead to lengthy processes and increased memory consumption when handling large data sets that exceed the capacity of the available RAM. In R, data structures that exceed $10 - 20\%$ of a computer's RAM are considered inefficient, while data sets exceeding 50% of available RAM are "essentially unusable" [14].

Consequently, data size and computational complexity are a major concern for analysts working with R. Given the widespread of large-scale data sources, addressing this issue becomes imperative. To improve memory usage and computation speed, various packages leveraging distinct techniques have been developed. The first objective is reached using fixed storage (disk space) in combination with RAM and this can be achieved using packages like `bigmemory` [14]. The second objective is reached using parallelization techniques that can be performed by various packages, such as `doParallel` [18].

The `bigmemory` package serves as a powerful tool for the efficient data management and manipulation within the R environment. While retaining the statistical functionalities of the R software, `bigmemory` is conceived to bridge the gap between R and programming languages like `C` and `C++` that excel in handling large datasets that may exceed the available RAM limits. It achieves its objective of minimizing the memory footprint and enhancing the computational performance using the `C++` programming language into R, thereby enabling the use of shared memory and file-backed data structures. This approach involves saving memory-mapped files, containing large-scale data, on disk.

Parallel computing refers to the partitioning of a problem into discrete instructions executed simultaneously by multiple processors, as opposed to the sequential execution by a single processor. With the exponential growth of data, the demand for accelerated computations has escalated. Increasing the frequency of processors ceased to be a feasible solution due to the subsequent increase in power consumption and heat generation. Hence, parallel computing systems were conceived as a strategic solution to overcome these constraints.

R was not originally conceived as a software environment for parallelization, but the community has addressed this necessity by developing multiple packages that facilitate the implementation of parallel procedures within R. Specifically, the `doParallel` package [18] stands out as the most comprehensive, combining the functionalities of the `foreach`, `iterators` and `parallel` packages.

3 Fuzzy Clustering Implementations for Big Data

All R packages where fuzzy clustering algorithms are implemented either rely on external `C++` code or on `for` loop iterations. The functions relying on `for` loops are highly inefficient, specially when working with large datasets. The initial approach to optimise computational time in the fuzzy clustering application within R consists in using the `foreach` loop in conjunction with iterator functions. This method was chosen to establish the groundwork for the subsequent

implementation of parallel computing and demonstrates its effectiveness even in the sequential application, outperforming the `for` loop.

The direct transition from the `for` loop to the `foreach` loop initially resulted in sub-optimal performances. To address this inefficiency, the `chunksize` argument of the `iter` function was implemented. This adaptation led to a reduced computing time when dividing the dataset into a few equal chunks.

The efficiency of `foreach` loop is evident when the instances are distributed into equally sized "sub-instances". Furthermore, a substantial reduction in computational time is achieved through the adoption of the `dist` function in the `proxy` package.

Significant improvements can be achieved through the use of parallel computing. The structure of this approach aligns perfectly with this context and, when synergistically integrated with other powerful tools, such as the `bigmemory` package, it holds the potential to significantly reduce computational times.

The use of parallel computing in this scenario is straightforward with the functionalities of the `foreach` package, which are already implemented to reduce computing time.

In the current computing landscape, where most machines operate on a single processor with multiple cores,[2] the `doParallel` package, coupled with the `cores` option, serves as the most fitting parallel back-end. Registering such a parallel back-end is accomplished using `registerDoParallel(cores = num_cores)` , where `num_cores` is the number of cores designated for code execution. It is worth noting that, while registering a parallel back-end is beneficial for large datasets on adequately equipped machines, it could drastically increase the execution time required to run the algorithm for a relatively small dataset or on less powerful machines. To address this, the sequential execution of the function for the fkm is made possible using an `if` statement. The user has control over the input argument `num_cores`, that enables a specific procedure based on the desired computational configuration: the user can specify to use a single core to execute the function with sequential computing, using `num_cores = 1` .

The integrated use of both parallel computing with `doParallel` and `bigmemory` in the computation of distance matrix, membership degrees and cluster centers is as follows:

```
1 distances <- proxy::dist(data, centers)[]
```

Code 11.1. Computation of distance matrix with dist

```
1 memberships <- foreach(d = iter(distances, by = "row",
```

[2] A processor is the electronic chip located in the computer which sends instructions to the core, the executing unit placed in the processor.

```
chunksize = size), .combine = "rbind") %dopar% {
    nr <- nrow(d)
    a <- matrix(0, nrow = nr, ncol = k)
    for (i in 1:nr) {
        for (j in 1:k) {
            a[i, j] <- 1 / (sum((d[i,j] / d[i,])^(2 / (m -
    1))))
        }
    }
    return(a)
}
memberships[is.na(memberships)] <- 1
```

Code 11.2. Parallel computation of membership degrees

```
new_centers <- foreach (i = iter(memberships, by = "col"),
.combine = "rbind", .packages = "bigmemory")%dopar%{
    attdata <- attach.big.matrix(datadesc)
    new_cnt <- (t(i^m) %*% attdata[]) / sum(i^m)
    return(new_cnt)
}
```

Code 11.3. Parallel computation of centers

It is important to note that the `chunksize` argument must be specified in a manner that ensures each parallel worker receives data of consistent sizes to enhance efficiency. Therefore, the proposed specified value for the variable `size` in the computation of membership degrees is `size = ceiling(nrow(data) / num_cores)` , ensuring the equal division of the data matrix for each parallel worker.

Additional auxiliary functionalities are requisite for the algorithm to converge successfully.

The implementation of the *fk*m method for big data is the function `bigFKM`:

```
>bigFKM(data, centers, m, num_cores = 4,
+    epsilon = 1e-06, max_iter = 1e+06, RS = 1)
```

Code 11.4. Call of bigFKM function

The input arguments are defined as:

– `data`: the dataset, stored in a matrix or `bigmatrix` object, devoid of `NA` values.

- `centers`: an integer value specifying the number of clusters, or a matrix of cluster centers used to initialize the algorithm. If not specified, the default number of clusters is set to 2.
- `m`: the fuzziness parameter. If not specified, the fuzziness parameter is set to 2 by default.
- `num_cores`: the number of cores available for parallel computing. If not specified, the default number of cores is set to 4.
- `epsilon`: a small positive numerical value used as a threshold for the convergence criteria. If not specified, ε is set to `1e-06` by default.
- `max_iter`: the number of maximum iteration for which the algorithm has to seek convergence. If not specified, the maximum number of iterations is set to `1e+06` by default.
- `RS`: the number of random starts the algorithm is tasked to perform. If not specified, the default number of random starts is set to 1.

In a similar way, variants of the f*k*m algorithm for big data have been implemented: `bigFKM.gk`, `bigFKM.ent` and `bigFKM.gk.ent`, among others.

4 Time Comparison

This section is devoted to a time comparison between the parallel and sequential implementations of the f*k*m algorithm. The data were randomly generated from normal distributions. We report the results obtained in case of $k = 3$ clusters for different sample sizes: 1000, 10000, 100000, 500000 and 1000000. For each sample size 500 replications were generated.

The comparisons are further controlled by using a fixed number of iterations and fixed convergence parameters, ensuring a consistent evaluation environment.

The simulation study involves the `FKM` function in the `fclust` package, the here-proposed `bigFKM` function in a sequential environment and in a parallel computing environment with 4 and 8 cores. The resulting average execution times are reported in Table 1 and are presented in Fig. 1.

Table 1. Average execution times (in seconds) for different sample sizes.

Sample size	FKM	bigFKM$_1$	bigFKM$_4$	bigFKM$_8$
1000	0.4865	1.6860	3.0723	4.0505
10000	8.0113	3.5985	4.0259	4.8751
100000	52.2411	29.0759	13.2243	12.5676
500000	268.6898	108.3539	52.3845	44.0424
1000000	696.2998	294.2948	125.3238	74,2141

A comprehensive analysis of the plot depicted in Fig. 1 reveals contrasting time complexities exhibited by the functions according to the size of the dataset.

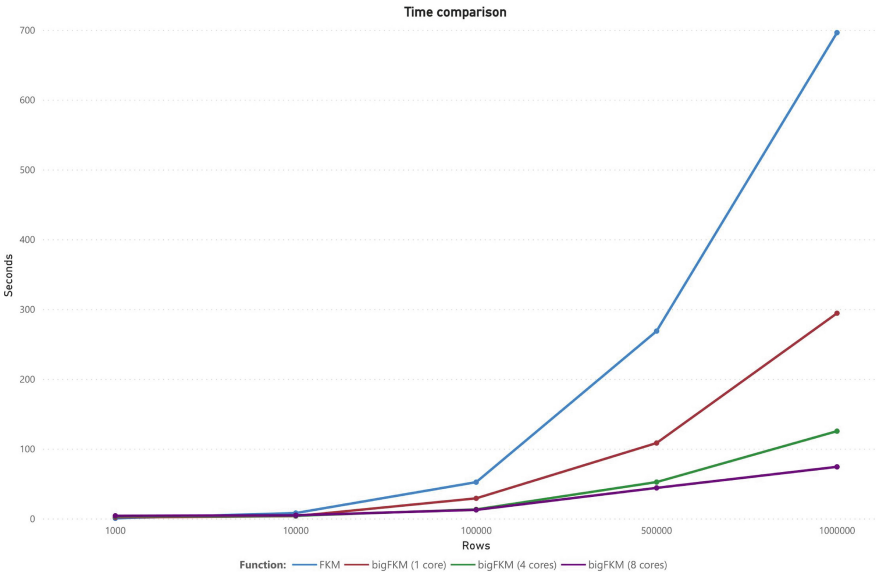


Fig. 1. Average execution times of `FKM` and `bigFKM` functions.

Notably, the initial scenario describes the inefficacy of applying parallel computing for relatively small datasets. This is corroborated by the observation that both the `FKM` and `bigFKM` functions achieve the shortest execution times using a single processing core. In this context, the communication overhead associated with parallel computing is greater than the execution time.

For a dataset size of 10000, the plot exposes the impact of time optimization changes within the `bigFKM` function. Here, the sequential execution of `bigFKM` outperforms all implementations, while the `FKM` function demonstrates the least favourable performance with the highest execution time.

Beyond the 10000 rows threshold, the influence of parallel computing becomes progressively evident at dataset sizes of 100000, 500000, and 1000000 rows. In all these scenarios, the parallel execution of the `bigFKM` function leveraging 8 cores, consistently emerges as the optimal solution. It is followed by the `bigFKM` function using 4 cores, the sequential execution of the `bigFKM` function, and lastly, the `FKM` function demonstrating a clear performance hierarchy according to the level of parallelism employed.

5 Concluding Remarks

`R` functions that facilitate the parallel implementation of fuzzy clustering algorithms, specifically designed for big data analysis, have been proposed. To overcome the limitations of existing functions in terms of execution time and memory consumption, this proposal exploits the capabilities of modern resources

such as the `bigmemory` [14] and the `doParallel` package [18] to enable parallel processing. In addition, a comparative analysis was conducted to evaluate the performance in terms of execution time. The results obtained illustrated the advantages of parallel processing over traditional sequential approaches. These new functions will be made available in an `R` package.

Acknowledgements. The research in this paper has been supported by the Sapienza University Grant "New developments on clustering of text, functional and other complex data" (Bando per la ricerca di Ateneo 2021).

References

1. Fauzi Bagus, A.F., Pramana, S.: `advclust`: object oriented advanced clustering. R package version 0.4 (2019). https://cran.r-project.org/web/packages/advclust
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithm. Plenum Press, New York (1981)
3. Bezdek, J.C., Hathaway, R.J., Huband, J.M., Leckie, C., Kotagiri, R.: Approximate clustering in very large relational data. Int. J. Intell. Syst. **21**, 817–841 (2006)
4. Cebeci, Z., Yildiz, F., Kavlak, A.T., Cebeci, C., Onder, H.: `ppclust`: probabilistic and possibilistic cluster analysis. R package version 0.1.3 (2019) https://CRAN.Rproject.org/package=ppclust
5. Ferraro, M.B.: Fuzzy k-Means: history and applications. Econom. Stat. **30**, 110–123 (2024)
6. Ferraro, M.B., Giordani, P.: A toolbox for fuzzy clustering using the R programming language. Fuzzy Sets Syst. **279**, 1–16 (2015)
7. Ferraro, M.B., Giordani, P., Serafini, A.: Fclust: an R package for fuzzy clustering. R J. **11**, 1–18 (2019)
8. Ferraro, M. B., Giordani, P. : A new fuzzy clustering algorithm with entropy regularization. In: 9th Meeting of the Classification and Data Analysis Group (Cladag 2013), pp. 195–198. CLEUP (2013)
9. Ghadiri, N., Ghaffari, M., Nikbakht, M.A.: BigFCM: fast, precise and scalable FCM on hadoop. Future Gener. Comput. Syst. **77**, 29–39 (2013)
10. Gustafson, D.E., Kessel, W.C.: Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of the 1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, pp. 761–766 (1979)
11. Hathaway, R.J., Bezdek, J.C.: Extending fuzzy and probabilistic clustering to very large data sets. Comput. Stat. Data Anal. **51**, 215–234 (2006)
12. Hidri, M.S., Zoghlami, M.A., Ayed, R.B.: Speeding up the large-scale consensus fuzzy clustering for handling Big Data. Fuzzy Sets Syst. **348**, 50–74 (2018)
13. Jha, P., Tiwari, A., Bharill, N., et al.: Scalable incremental fuzzy consensus clustering algorithm for handling big data. Soft. Comput. **25**, 8703–8719 (2021)
14. Kane, M., Emerson, J.W., Weston, S.: Scalable strategies for computing with massive data. J. Stat. Softw. **55**, 1–19 (2013)
15. Li, R.-P., Mukaidono, M.: Gaussian clustering method based on maximum-fuzzy-entropy interpretation. Fuzzy Sets Syst. **102**, 253–258 (1999)
16. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., Lin, C.-C.: `e1071`: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-2 (2019) https://cran.r-project.org/web/packages/e1071

17. Venkat, R., Reddy, K.S. (2019). Dealing big data using fuzzy C-Means (FCM) clustering and optimizing with gravitational search algorithm (GSA). In: 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, pp. 465–467 (2019)
18. Folashade, D., Weston, S., Tenenbaum, D.: `doParallel`: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.17 (2022) https://cran.r-project.org/web/packages/doParallel

On Topologically Typical Bivariate Extreme Value Copulas

Nicolas Pascal Dietrich(✉) and Wolfgang Trutschnig

Department for Artificial Intelligence and Human Interfaces, University of Salzburg, Hellbrunnerstrasse 34, Salzburg 5020, Austria
nicolaspascal.dietrich@plus.ac.at, wolfgang@trutschnig.net

Abstract. Motivated by the fact that (in the sense of Baire categories and working with the uniform distance d_∞) typical bivariate copulas are completely dependent, we prove that even the subclass of all mutually completely dependent copulas with full support are typical (co-meager). Additionally, considering the subclass of Extreme Value copulas, working with so called Pickands dependence measures, i.e., univariate probability measures with expected value $\frac{1}{2}$, allows not only to determine the support of the Extreme Value copula via the afore-mentioned measure, but also, working with Markov-kernels, to characterize the discrete component of the Extreme Value copula by the point masses of the Pickands dependence measure. Establishing a homeomorphism between the spaces of all Pickands dependence measures and the family of Extreme Value copulas and using Markov-kernels, we derive the surprising result that, in contrast to the family of all bivariate copulas, a typical Extreme Value copula has degenerated discrete component, is not absolutely continuous and has full support.

Keywords: Extreme Value copula · Markov kernel · Lebesgue decomposition · Baire-category

1 Introduction

Forming the building blocks between multivariate distribution functions and their univariate marginal distributions (see [25]) and their resulting importance in the context of modeling stochastic dependence (see [6,10–12,18,20,23]), throughout the last decades copulas became a prominent tool in statistics and probability theory. Equipping the family of all copulas with the standard uniform metric d_∞ and viewing it from a purely analytical/topological standpoint, the increasing interest in copulas gave rise to numerous dense sub-families (e.g. Bernstein copulas, shuffels, checkerboards) of copulas, see e.g. [6,20]. However, going even further, the question arises, whether such sub-families are 'small'

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 102–109, 2024.
https://doi.org/10.1007/978-3-031-65993-5_12

or 'large' in some sense, i.e., whether they cover a large enough portion of the space of copulas. Working with Baire categories (see e.g. [21]), topology offers a natural way to differentiate between 'small' and 'large' sets. More precisely, we call a given subset of a topological space $(\mathcal{T}, \tau)$ *nowhere dense* if, and only if the interior of the closure of that set is empty in $(\mathcal{T}, \tau)$. Moreover, we call a set in $(\mathcal{T}, \tau)$ *meager* (or, equivalently of *first Baire-category*), if it is the countable union of nowhere dense sets. A set is of *second Baire-category*, if it is not meager and, finally, a set is called *co-meager*, if it is the complement of a meager set. Proceeding as in [2] and returning to the concept of 'small' and 'large' sets, in a complete metric space meager sets are the 'small' sets, sets of second category are interpreted as 'not small' and co-meager sets are the 'large' sets. Moreover, having a topological space $(\mathcal{T}, \tau)$, we call elements of a co-meager set a *typical* element and elements of a meager set an *atypical* element of that space. For a variety of manuscripts considering copulas in the context of Baire categories, we refer to [3–5,8].

Working with the afore-mentioned concepts the authors in [8] showed that a typical bivariate copula is singular in the sense that its associated doubly stochastic measure is fully supported on a set of Lebesgue-measure zero. In fact, as proved by C.W. Kim (see [14]) even the set of all completely dependent copulas is co-meager in the space of all copulas (all w.r.t. d_∞). Building upon this result, as discussed in Sect. 3, an even stronger result is proved: A typical copula is mutually completely dependent and its associated doubly stochastic measure has full support.

Due to their simple algebraic form and due to several practical applications in finance and hydrology (see [19,24]) probably one of the most prominent subclass of copulas is the family of Extreme Value copulas. Working with Extreme Value copulas in Sect. 4, we establish that they can be fully characterized by a univariate probability measure, a so called Pickands dependence measure, on the unit interval with expectation $\frac{1}{2}$. Note that characterizing Extreme Value copulas via such measures is not new, see [1,7,9,26]. Building upon the afore-mentioned characterization and simply working with univariate probability measures we are, however, able to simplify some of the results given in [17,28]. It becomes apparent (see Sect. 4) that we are not only capable of fully determining the support of an Extreme Value copula via its associated univariate Pickands dependence measure, but that even the discrete component of an Extreme Value copula can be characterized solely by the point masses of its associated Pickands dependence measure. Moreover, it is well known (see [9,20,22]) that Extreme Value copulas can also be represented via so-called Pickands-dependence functions. Thus, using the afore-mentioned characterizations, we establish that the spaces of Pickands dependence measures, the family of Extreme Value copulas, as well as the space of Pickands dependence functions are homeomorphic, which, focusing on the main part of Sect. 4 and keeping in mind that in the whole family of copulas a typical copula is purely discrete (in the sense that Lebesgue-almost all conditional distribution functions are discrete), yields the surprising result that typical Extreme Value copulas have degenerated discrete component. Building upon that, we can even show a more striking result: A typical Extreme Value copula has full support, non-degenerated discrete component

and is not absolutely continuous. Due to space limitations, proves are provided in the journal version of this paper.

2 Notation and Preliminaries

In the sequel $\mathcal{C}$ will denote the family of all bivariate copulas. Considering a copula $C \in \mathcal{C}$, as usual, the associated doubly stochastic measure μ_C, is given by $\mu_C([0,x] \times [0,y]) := C(x,y)$ for all $x, y \in \mathbb{I}$ with and $\mathbb{I} := [0,1]$. For an arbitrary metric space (S,d) we denote the Borel σ-field on S by $\mathcal{B}(S)$. Moreover, for every measure ν on $\mathcal{B}(S)$ the support of ν, i.e., the complement of the union of all open sets U with the property $\nu(U) = 0$, will be denoted by $\mathrm{supp}(\nu)$. Throughout this contribution, the support of a copula $C \in \mathcal{C}$ will be the support of its corresponding doubly stochastic measure μ_C. Considering the uniform metric d_∞ on $\mathcal{C}$ it is well-known that $(\mathcal{C}, d_\infty)$ is a compact metric space. For more background on copulas and doubly stochastic measures we refer to [6,20]. Furthermore, $\mathcal{P}(S)$ denotes the family of all probability measures on $\mathcal{B}(S)$, δ_x denotes the Dirac-measure in $x \in S$ and λ_2 will be the 2-dimensional Lebesgue measure on $\mathcal{B}(\mathbb{I}^2)$.

Working with conditional distribution functions of copulas, throughout this contribution, Markov kernels will be an essential tool. To be more precise, a *Markov kernel* from $\mathbb{I}$ to $\mathbb{I}$ is a map $K : \mathbb{I} \times \mathcal{B}(\mathbb{I}) \to \mathbb{I}$ fulfilling that for every fixed $F \in \mathcal{B}(\mathbb{I})$ the function $x \mapsto K(x,F)$ is Borel-measurable and for all $x \in \mathbb{I}$ the map $F \mapsto K(x,F)$ is a probability measure on $\mathcal{B}(\mathbb{I})$. It is well known (see e.g. [27]) that every copula $C \in \mathcal{C}$ induces a Markov kernel which will be denoted by K_C. Fixing $x \in \mathbb{I}$ and defining the x-section G_x of an arbitrary set $G \in \mathcal{B}(\mathbb{I}^2)$ by $G_x := \{y \in \mathbb{I} : (x,y) \in G\} \in \mathcal{B}(\mathbb{I})$ the *disintegration theorem* (see [13, Sect. 5] and [15, Sect. 8]) yields that

$$\mu_C(G) = \int_{\mathbb{I}} K_C(x, G_x)\, \mathrm{d}\lambda(x). \tag{1}$$

In what follows a measure ν on $\mathcal{B}(\mathbb{I}^2)$ will be called singular if, and only if there exists some set $G \in \mathcal{B}(\mathbb{I}^2)$ with $\lambda_2(G) = 0$, $\nu(G) = \nu(\mathbb{I}^2)$ and additionally ν does not have any point-masses. Observing that the doubly stochastic measure μ_C associated with a copula $C \in \mathcal{C}$ always has degenerated discrete component, defining the absolutely continuous, discrete and singular component of a copula, in the sense of the Lebesgue-decomposition theorem of μ_C, turns out to be a delicate issue. Following [28] and using the Lebesgue-decomposition of the Markov kernel K_C into purely absolutely continuous, discrete and singular sub-kernels K_C^{abs}, K_C^{dis} and K_C^{sing}, respectively, i.e.,

$$K_C(x,F) = K_C^{abs}(x,F) + K_C^{dis}(x,F) + K_C^{sing}(x,F) \tag{2}$$

for $x \in \mathbb{I}$ and $F \in \mathcal{B}(\mathbb{I})$, see [16], allows for a very natural definition of the absolutely continuous, discrete and singular component of μ_C. To be more precise: Considering a copula $C \in \mathcal{C}$, motivated by the disintegration theorem and

Eq. (2), we define the absolutely continuous, discrete and singular component of μ_C by

$$\mu_C^j(E \times F) := \int_E K_C^j(\mathbf{x}, F)\mathrm{d}\lambda(x) \tag{3}$$

for $j \in \{abs, dis, sing\}$ and $E, F \in \mathcal{B}(\mathbb{I})$.
At last, we call a map $h\colon \mathbb{I} \to \mathbb{I}$ λ-preserving if, and only if $\lambda^h(F) = \lambda(F)$ for every $F \in \mathcal{B}(\mathbb{I})$ and a copula $C \in \mathcal{C}$ completely dependent if, and only if its mass is concentrated on the graph of h. If, in addition, h is bijective, we call the associated completely dependent copula mutually completely dependent. Throughout this contribution we denote the space of all completely dependent copulas by $\mathcal{C}_{cd}$ and the space of all mutually completely dependent copulas by $\mathcal{C}_{mcd}$. For more information on completely dependent copulas we refer to e.g. [20].

3 Baire-Category Results for the Whole Family $\mathcal{C}$

Considering the whole space of copulas, C.W. Kim came up with the result (see [14]) that typical copulas are completely dependent, which is surprising in so far as these kind of copulas seem very pathological and irregular.

Theorem 1. *([14]) $\mathcal{C}_{cd}$ is co-meager in $(\mathcal{C}, d_\infty)$.*

Building upon the afore-mentioned theorem, we are able to prove the even stronger result that typical copulas are mutually completely dependent.

Corollary 1. *$\mathcal{C}_{mcd}$ is co-meager in $(\mathcal{C}, d_\infty)$.*

Viewing copulas as doubly stochastic measures, a typical copula has full support.

Lemma 1. *The family $\{C \in \mathcal{C}\colon \mathrm{supp}(\mu_C) = \mathbb{I}^2\}$ is co-meager in $(\mathcal{C}, d_\infty)$.*

Altogether, using the fact that finite intersections of co-meager sets are again co-meager (see [21]) we obtain the following corollary.

Corollary 2. *A typical copula in $(\mathcal{C}, d_\infty)$ is mutually completely dependent and has full support.*

4 Category Results for Extreme Value Copulas

A copula $C \in \mathcal{C}$ is an Extreme Value copula (EVC) if, and only if there exists a copula $B \in \mathcal{C}$, such that

$$C(x, y) = \lim_{n\to\infty} B^n(x^{\frac{1}{n}}, y^{\frac{1}{n}})$$

holds for all $x, y \in \mathbb{I}$. Throughout this section, the space of all Extreme Value copulas will be denoted by $\mathcal{C}_{ev}$. Working with the family of all Pickands dependence functions on $\mathbb{I}$, i.e.,

$$\mathcal{A} := \{A\colon \mathbb{I} \to \mathbb{I}\colon A \text{ is convex, } \max\{1-x, x\} \le A(x) \le 1 \text{ and } A(0) = 1 = A(1)\},$$

according to [20,22] a copula C_A is an EVC if, and only if there exists $A \in \mathcal{A}$ such that

$$C_A(x,y) := (xy)^{A\left(\frac{\log(x)}{\log(xy)}\right)}, \tag{4}$$

for every $x, y \in (0,1)$. Defining the space of all Pickands dependence measures by

$$\mathcal{P}_{\mathcal{A}} := \left\{ \vartheta \in \mathcal{P}(\mathbb{I}) \colon \int_{\mathbb{I}} x \mathrm{d}\vartheta(x) = \frac{1}{2} \right\},$$

it is known (see [28]) that every $\vartheta \in \mathcal{P}_{\mathcal{A}}$ induces a unique Pickands dependence function and vice versa. Put differently, following [1,7,9,26], there exists a map $\Upsilon \colon (\mathcal{P}_{\mathcal{A}}, \tau_w) \to (\mathcal{A}, \|\cdot\|_\infty)$ linking $A \in \mathcal{A}$ and $\vartheta \in \mathcal{P}_{\mathcal{A}}$ via

$$(A(t) =)\, \Upsilon(\vartheta)(t) := 1 - t + 2 \int_{[0,t]} \vartheta([0,z]) \mathrm{d}\lambda(z) \tag{5}$$

for all $\vartheta \in \mathcal{P}_{\mathcal{A}}$ and τ_w denotes the topology induced by weak convergence and $\|\cdot\|_\infty$ denotes the supremum-norm. Considering the one-to-one relation between Pickands dependence functions and EVC, there exists a map $\Phi \colon (\mathcal{A}, \|\cdot\|_\infty) \to (\mathcal{C}_{ev}, d_\infty)$ defined via Eq. (4). To simplify notation, we introduce the function $F_A \colon \mathbb{I} \to \mathbb{I}$, defined by $F_A(t) := A(t) + D^+A(t)(1-t)$ for $t \in [0,1)$ and $F_A(1) := 1$, whereby D^+A denotes the right-hand derivative of A. Considering conditional distributions of EVC, according to [28] a version of the Markov-kernel of the EVC C_A is given by

$$K_A(x,[0,y]) := \begin{cases} 1, & \text{if } x \in \{0,1\} \\ \frac{C_A(x,y)}{x} F_A\left(\frac{\log(x)}{\log(xy)}\right), & \text{if } x,y \in (0,1) \\ y, & \text{if } (x,y) \in (0,1) \times \{0,1\} \end{cases} \tag{6}$$

for $x, y \in \mathbb{I}$ and $A \in \mathcal{A}$.

Working with the afore-mentioned one-to-one relation between Pickands dependence functions $A \in \mathcal{A}$, Pickands dependence measures $\vartheta \in \mathcal{P}_{\mathcal{A}}$ and EVCs $C \in \mathcal{C}_{ev}$, respectively, it is straightforward to show that the maps Φ and Υ are homeomorphisms.

Lemma 2. *The maps Φ, Υ and $\Phi \circ \Upsilon$ are homeomorphisms.*

4.1 Category Results

Fixing $t \in (0,1)$ and defining the function $f^t(x) := x^{\frac{1}{t}-1}$ for $x \in \mathbb{I}$, it is known (see [28]) that the mass of the discrete component of an EVC (if any) is concentrated on the graph of these functions. Having a Pickands dependence measure $\vartheta \in \mathcal{P}_{\mathcal{A}}$, we define

$$L = sup\{x \in \mathbb{I} \colon \vartheta([0,x]) = 0\}, \quad R = inf\{x \in \mathbb{I} \colon \vartheta([0,x]) = 1\}, \tag{7}$$

where the supremum of the empty set is interpreted as zero. Following [28], L and R can also be defined working with Pickands dependence functions. Remember,

see [28], that the support of an extreme value copula $C \in \mathcal{C}_{ev}$ can be fully determined by the functions f^L and f^R, respectively, i.e.,

$$\operatorname{supp}(\mu_C) = \{(x,y) \in \mathbb{I}^2 \colon f^L(x) \leq y \leq f^R(x)\}. \tag{8}$$

Considering the function f^t, an EVC carries mass on the graph of f^t if, and only if its associated probability measure $\vartheta \in \mathcal{P}_{\mathcal{A}}$ has an atom in t. In fact, the next lemma shows that the mass of the graph corresponds to the mass of ϑ.

Theorem 2. *Let $C \in \mathcal{C}_{ev}$ be an EVC with associated Pickands dependence measure $\vartheta \in \mathcal{P}_{\mathcal{A}}$. Then ϑ has an atom in $(0,1)$ if, and only if $\mu_C^{dis}(\mathbb{I}^2) > 0$. Moreover, the identity*

$$\mu_C(\Gamma(f^t)) = \frac{2t(1-t)}{A(t)}\vartheta(\{t\}) \tag{9}$$

holds for all $t \in (0,1)$.

Considering Eq. (8), an Extreme Value copula has full support if, and only if $L = 0$ and $R = 1$. In fact, applying Lemma 2 together with the afore-mentioned fact, yields that a typical EVC has full support.

Corollary 3. *The set $\{C \in \mathcal{C}_{ev} \colon \operatorname{supp}(\mu_C) = \mathbb{I}^2\}$ is co-meager in $(\mathcal{C}_{ev}, d_\infty)$.*

Working with Lemma 2 we obtain the main theorem of this contribution as direct consequence of the fact that typical Pickands dependence functions are everywhere differentiable, typical Pickands dependence measures are not absolutely continuous and the fact that the regularity of the Pickands dependence measure and Pickands dependence function respectively propagates to the EVC and vice versa. Contrary to the family of all copulas, the next theorem states that a typical EVC has degenerated discrete component.

Theorem 3. *The set $\{C \in \mathcal{C}_{ev} \colon \mu_C^{dis}(\mathbb{I}^2) = 0\} \cap \{C \in \mathcal{C}_{ev} \colon C$ is not abs.cont.$\}$ is co-meager in $\mathcal{C}_{ev}$.*

Combining Theorem 3 and Lemma 3, yields the final result of this manuscript:

Corollary 4. *A typical EVC has degenerated discrete component, is not absolutely continuous and has full support.*

Illustrating the afore-mentioned results, we finally provide an example of a non-typical EVC.

Example 1. Defining the probability measure ϑ by

$$\vartheta := \frac{1}{5}\delta_{\frac{1}{4}} + \frac{3}{5}\delta_{\frac{1}{2}} + \frac{1}{5}\delta_{\frac{3}{4}},$$

clearly $\vartheta \in \mathcal{P}_{\mathcal{A}}$. Thus, working with the formula given in Eq. (5), we obtain the Pickands-dependence function

$$A(t) := \begin{cases} 1-t, & \text{if } t \in [0, \frac{1}{4}) \\ -\frac{3}{5}t + \frac{9}{10}, & \text{if } t \in [\frac{1}{4}, \frac{1}{2}) \\ \frac{3}{5}t + \frac{3}{10}, & \text{if } t \in [\frac{1}{2}, \frac{3}{4}) \\ t, & \text{if } t \in [\frac{3}{4}, 1]. \end{cases}$$

According to Corollary 4, C_A is a non-typical EVC, since it has a discrete component which, considering Theorem 2, is concentrated on the graphs of the functions $f^{\frac{1}{4}}$, $f^{\frac{1}{2}}$ and $f^{\frac{3}{4}}$, respectively. Moreover, C_A does not have full support, since according to Eq. (8) its support is given by $\text{supp}(\mu_{C_A}) = \{(x,y) \in \mathbb{I}^2 \colon f^{\frac{1}{4}}(x) \leq y \leq f^{\frac{3}{4}}(x)\}$. Figure 1 depicts a sample of the copula C_A.

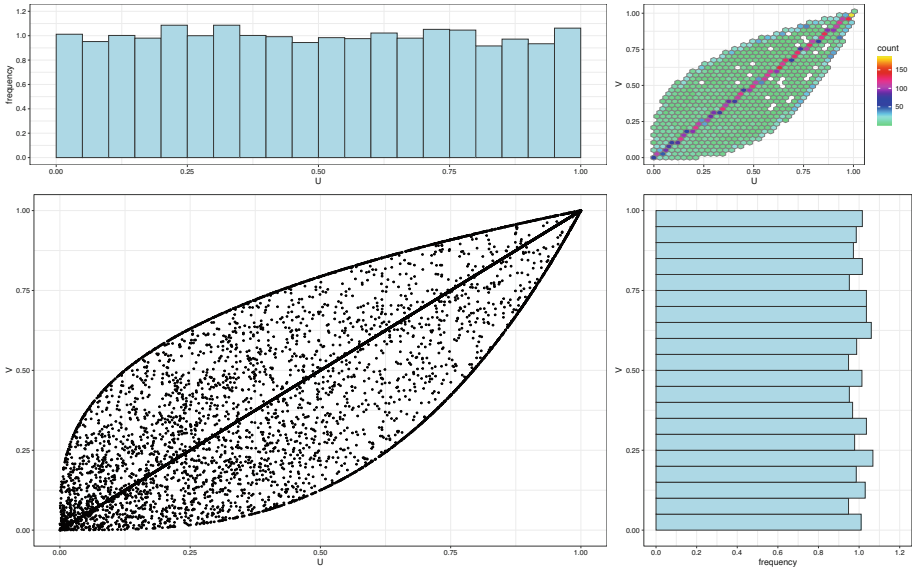


Fig. 1. Sample of size 10000 of the EVC C_A, where A is the Pickands dependence function considered in Example 1, its histogram and the two marginal histograms. The sample has been generated via conditional inverse sampling.

Acknowledgements. The first author gratefully acknowledges the financial support from Red Bull GmbH within the ongoing Data Science collaboration with the university of Salzburg.

References

1. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of extremes: Theory and Applications. Wiley Series in Probability and Statistics. Wiley, Chichester (2004)
2. Bruckner, A.M., Thomson, B.S.: Real Analysis. Prentice-Hall, London (1997)
3. Durante, F., Fernández-Sánchez, J., Ignazzi, C.: Baire category results for stochastic orders. Rev. Real Acad. Cienc. Exactas Fis. Nat. - A: Mat. **116**, 188 (2022)
4. Durante, F., Fernández-Sánchez, J., Trutschnig, W.: Baire category results for exchangeable copulas. Fuzzy Sets Syst. **284**, 146–151 (2016)

5. Durante, F., Fernández-Sánchez, J., Trutschnig, W.: Baire category results for quasi-copulas. Depend. Model. **4**, 215–223 (2016)
6. Durante, F., Sempi, C.: Principles of Copula Theory. Chapman and Hall/CRC (2015)
7. Falk, M., Häsler, J., Reiss, R.D.: Laws of Small Numbers: Extremes and Rare Events. Springer Basel (2011)
8. Fernández-Sánchez, J., Trutschnig, W.: A typical copula is singular. J. Math. Anal. Appl. **430**, 517–527 (2015)
9. Limit theory for multivariate sample extremes: Haan, L.d., Resnick, S.I. Probab. Theory Relat. Fields **40**, 317–337 (1977)
10. Jaworski, P., Durante, F., Härdle, W.K.: Copulae in Mathematical and Quantitative Finance. In: Lecture Notes in Statistics—Proceedings, Vol. 213. Springer, Berlin Heidelberg (2013)
11. Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T., (eds.): Copula Theory and its Applications. In: Lecture Notes in Statistics—Proceedings, Vol. 198. Springer, Berlin Heidelberg (2010)
12. Joe, H.: Dependence Modeling with Copulas. Chapman and Hall/CRC, London (2014)
13. Kallenberg, O.: Foundations of Modern Probability. Springer, New York (2002)
14. Kim, C.: Uniform approximation of doubly stochastic operators. Pac. J. Math. **26**, 515–527 (1968)
15. Klenke, A.: Wahrscheinlichkeitstheorie. Springer Lehrbuch Masterclass Series, Berlin Heidelberg (2008)
16. Lange, K.: Decompositions of substochastic transition functions. Proc. Amer. Math. Soc. **37**, 575–580 (1973)
17. Mai, J.F., Scherer, M.: Bivariate extreme-value copulas with discrete Pickands dependence measure. Extremes **14**, 311–324 (2011)
18. Mai, J.-F., Scherer, M.: Simulating copulas: stochastic models, sampling algorithms, and applications. In: Series in Quantitative Finance, Vol. 4. Imperial College Press, London (2012)
19. McNeil, A., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, New Jersey (2005)
20. Nelsen, R.B.: An introduction to copulas. Springer Series in Statistics, 2nd ed. Springer, New York (2006)
21. Oxtoby, J.C.: Measure and category. A survey of the analogies between topological and measure spaces, 2nd ed. Springer-Verlag (1980)
22. Pickands, J.: Multivariate extreme value distributions. In: Proceedings 43rd Session International Statistical Institute 2, pp. 859-878 (1981)
23. Rüschendorf, L.: Mathematical Risk Analysis. Springer Series in Operations Research and Financial Engineering, Springer, Heidelberg (2013)
24. Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R.: Extremes in Nature—An Approach Using Copulas. Springer, Dordrecht (2007)
25. Sklar, A.: Fonctions de répartition à N dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
26. Tawn, J.A.: Bivariate extreme value theory: models and estimation. Biometrika **75**, 397–415 (1988)
27. Trutschnig, W.: On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. **384**, 690–705 (2011)
28. Trutschnig, W., Schreyer, M., Fernández Sánchez, J.: Mass distributions of two-dimensional extreme-value copulas and related results. Extremes **19**, 405–427 (2016)

Semi-supervised Learning Guided by the Generalized Bayes Rule Under Soft Revision

Stefan Dietrich(✉), Julian Rodemann, and Christoph Jansen

Ludwig—Maximilians—Universität (LMU), Department of Statistics, Munich, Germany

dietrich.stefan@campus.lmu.de, rodemann@stat.uni-muenchen.de, christoph.jansen@stat.uni-muenchen.de

Abstract. We provide a theoretical and computational investigation of the Gamma-Maximin method with soft revision, which was recently proposed as a robust criterion for pseudo-label selection (PLS) in semi-supervised learning. Opposed to traditional methods for PLS we use credal sets of priors ("generalized Bayes") to represent the epistemic modeling uncertainty. These latter are then updated by the Gamma-Maximin method with soft revision. We eventually select pseudo-labeled data that are most likely in light of the least favorable distribution from the so updated credal set. We formalize the task of finding optimal pseudo-labeled data w.r.t. the Gamma-Maximin method with soft revision as an optimization problem. A concrete implementation for the class of logistic models then allows us to compare the predictive power of the method with competing approaches. It is observed that the Gamma-Maximin method with soft revision can achieve very promising results, especially when the proportion of labeled data is low.

Keywords: Semi-supervised learning · Self-training · Pseudo-labeling · Generalized bayes rule · Imprecise probabilities · Credal sets · Soft revision · Classification · Machine learning

1 Introduction

In[1] this paper, we investigate Gamma-Maximin with soft revision for robust selection of pseudo-labeled data in semi-supervised learning (SSL) problems, as recently proposed in [17]. We use this introduction to collect those concepts from SSL that are most relevant to our article. For an in-depth motivation and extensive hints to relevant literature, we refer the reader to [17, Sect. 1].

Semi-Supervised Learning (SSL): The fundamental idea of SSL is to use both labeled data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ and unlabeled data $\mathcal{U} = \{(x_i, \mathcal{Y})\}_{i=n+1}^{m}$ for

[1] This work is partly based on Stefan Dietrich's Bachelor thesis "α-cuts as robust method of pseudo label selection within the context of semi-supervised learning".

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 110–117, 2024.
https://doi.org/10.1007/978-3-031-65993-5_13

training a classification model. In this approach, a limited set of labeled data points is combined with an often considerably larger set of unlabeled data points $m >> n$ to train a model $g(x) = \hat{y} \in \mathcal{Y}$ [7]. The incorporation of unlabeled data when training a model can intuitively be useful only if the fundamental SSL assumption is met, i.e., if the unlabeled data contains information about the relationship between the input data and the label of the assigned class [22]. In this paper, we will address *classification problems* with self-training methods and incremental addition of pseudo-labeled data as categorized by [19]. A self-training algorithm will, in the first phase, estimate a model from the labeled training data using supervised learning. Based on this model, labels are then predicted for the unlabeled data. These labels are referred to as *pseudo-labels*. In the second phase, the training data is enlarged by adding a selected pseudo-labeled data point according to some criterion. The same point is then removed from the unlabeled dataset, resulting in a new unlabeled dataset. This process is repeated until a stopping criterion is met. This paper will focus on the selection of a potentially favorable pseudo-labeled data point. That is, we take the predicted pseudo-labels as given and only deal with selecting one pseudo-labeled data point from the set of pseudo-labeled data.

Bayesian Analysis: In the entire process of PLS we encounter several sources of uncertainty. Traditionally, only the model used for predicting pseudo labels is utilized. Thus, the selection is performed based on the estimated model parameters, without considering the uncertainty associated with those. In the endeavor to establish a selection criterion that is robust against the initial model, i.e., more robust against the choice of model parameters, but at the same time efficiently utilizes the information from labeled data, the choice of a Bayesian approach appears fruitful [16]. Here, it is possible to decompose epistemic uncertainty into statistical uncertainty due to a lack of data and modeling uncertainty due to parametric assumptions. The fundamental idea of Bayesian pseudo-label selection (BPLS) is to choose an instance with a pseudo-label that is probable given the training data but not necessarily probable concerning the estimated model parameters. In the spirit of [15,16], we formalize PLS as a decision problem. The utility of a selected data point $(x_i, \mathcal{Y})$ shall now be the likelihood of being generated jointly with $\mathcal{D}$ by a model with parameters $\theta \in \Theta$ if we include it with the pseudo-label $\hat{y}_i \in \mathcal{Y}$ (obtained through any predictive model) in $\mathcal{D} \cup (x_i, \hat{y}_i)$.

Definition 1. *Let $(x_i, \mathcal{Y})$ be any decision (selection) from $\mathcal{U}$. We assign utility to each $(x_i, \mathcal{Y})$ given $\mathcal{D}$ and the pseudo-label $\hat{y} \in \mathcal{Y}$ by the following measurable utility function which is said to be the pseudo-label likelihood.*

$$u : \mathcal{U} \times \Theta \to \mathbb{R}\ ,\ ((x_i, \mathcal{Y}), \theta) \mapsto p(\mathcal{D} \cup (x_i, \hat{y}_i)|\theta) \tag{1}$$

The function u is a natural choice to assign utilities to pseudo-labels given the predicted pseudo-labels. Moreover, under a prior π, it can be shown that maximizing expected utility, i.e., applying the standard Bayes-criterion $\Phi(\cdot, \pi) : \mathcal{U} \to \mathbb{R};\ a \mapsto \Phi(a, \pi)) = \mathbb{E}_\pi(u(a, \cdot))$, with respect to u and π, corresponds to maximizing the pseudo marginal likelihood $p(\mathcal{D} \cup (x_i, \hat{y}_i))$ [16]. For any prior, the

action which results in the maximal marginal likelihood is Bayes optimal. If we now update the prior with data to a posterior, the standard Bayes criterion corresponds to the pseudo posterior predictive (PPP), i.e., the posterior predictive distribution of the predicted pseudo-labels.

Uncertainty in BPLS: An important source of uncertainty is the choice of the model, i.e., the distributional assumption under which the likelihood is calculated. So far, this distributional assumption is the same as the one used for pseudo-label prediction. For addressing uncertainty regarding model choice, a multi-model approach can be applied, which computes a weighted sum of likelihoods of multiple models, see [17]. Additionally, in the same paper, uncertainty regarding the prior selection was briefly discussed. The uncertainty associated with the prior will be thoroughly examined in the remainder of the paper through an implementation via an approximation and experiments.

2 Related Work

The robustness of pseudo-label selection has been extensively debated in the semi-supervised learning literature. Aminian et al. [1] propose an information theory based PLS approach adaptable to distributional shifts. Vandewalle et al. [21] aim to enhance model robustness against presumptive biases by facilitating model choice via the deviance information criterion within a semi-supervised learning framework. Rizve et al. [14] introduce an uncertainty-aware pseudo-label selection technique that competes with popular semi-supervised learning (SSL) methods anchored in consistency regularization. This technique involves selecting pseudo-labels by assessing both their probability scores and predictive uncertainty, the latter gauged by the prediction variance. Principled approaches, however, that hedge against misspecification of prior assumptions on which the PLS is based upon, are lacking so far. In the context of self-supervised learning, there is the notable exception of Lienen et al. [11]. Contrary to our approach, they use credal sets for labeling directly rather than for selecting pseudo-labels. Similar to us, they minimize the reliance on distributional assumptions. Finally, algorithms for computing optimal Γ-maximin decisions have also been discussed, e.g., in [9,20] in the context of decision making under ambiguity or in [12] in the context of robust Bayesian statistics. However, these approaches are not (directly) applicable here, as they assume finite state spaces or strictly restricted action spaces of their underlying decision problems, respectively.

3 Generalizied Bayesian PLS Under Soft Revision

In real-world applications, the unique choice of an adequate prior as required for Bayesian PLS is often not straightforward. Instead, several different candidate priors occur equally plausible. In order to make the selection of the pseudo-labeled data more robust against a potentially bad choice of prior in such situations, *credal sets* can be used. A credal set Π, as first introduced in [10] and heavily been applied in the imprecise probability (IP) community since then,

is simply a set of probability measures. Each element $\pi \in \Pi$ is interpreted as an equally plausible candidate prior for the application at hand. Consequently, in the case of modeling with credal sets, any method for pseudo label selection should also include the entire credal set instead of artificially reducing uncertainty to classical probability.

The transition from a precise prior to a credal set (also referred to as "generalized Bayes" [3]) naturally also requires a decision criterion that can process this information. We choose the Gamma-Maximin criterion, whose fundamental idea is to compute the expected utility under a prior distribution which contrasts the likelihood of the estimated model the most. Then, the instance corresponding to the highest expected utility with respect to its *least favorable prior* should be chosen. Even though Gamma-Maximin can be critized for being rather conservative, one can also argue that this very conservatism is often preferable as it safeguards against taking risky actions on the basis of unjustified assumptions [3]. In our case, it is better to choose a prior that may not align well with the likelihood rather than selecting one that aligns too closely with the likelihood.

The Gamma-maximin method in turn heavily depends on the choice of the credal set. The larger it is, the higher is the tendency to make a decision with complete disregard of the likelihood. On the other hand, a small credal set might not be expressive enough to sufficiently attenuate the likelihood. Therefore, in the following, we aim to restrict the set of all priors over which we calculate an infimum dynamically, i.e., for each iteration separately [17]. We adopt a credal set updating rule based on Cattaneo's α-cut rule with $\alpha \in (0, 1)$, also known as "soft revision" [2,6]. This updating strategy aims to update only those priors whose marginal likelihood is greater than or equal to α times the corresponding maximum marginal likelihood. Let, therefore, for the likelihood $\mathcal{L}$ and all $\pi \in \Pi$ be $m(\pi) = \int \mathcal{L}(\theta) f_\pi(\theta)\, d\theta$ the marginal likelihood, where $f_\pi(\theta)$ is the density of π with respect to the Lebesgue measure. Then $\Pi_\alpha = \{\pi \in \Pi \mid m(\pi) \geq \alpha \cdot \sup_\pi m(\pi)\}$ is the credal set of priors that we update to posteriors. The Gamma-Maximin decision criterion with soft revision is now defined as follows:

$$\Phi_\alpha : \boldsymbol{\mathcal{U}} \to \mathbb{R} \quad , \quad ((x_i, \boldsymbol{y})_i) \mapsto \inf_{\pi \in \Pi_\alpha} \mathbb{E}_\pi(u(a, \cdot)). \tag{2}$$

Φ_α addresses uncertainty in a general way without completely ignoring the likelihood. The steeper the likelihood, the smaller the credal set at a constant α.

To technically implement[2] the Gamma-Maximin criterion under soft revision, we first establish a model and a distribution assumption.[3] Here, we will consider datasets with bivariate labels and perform logistic regression on them. The labels are assumed to follow a Bernoulli distribution $Y_i \sim B(p_i)$, where p_i represents the response function that connects the linear predictor to the expected value or the parameters of the distributional assumption: $\mathbb{E}[Y_i] = p_i = \exp(\eta_i)/(1 + \exp(\eta_i))$.

[2] Implementation of the proposed method and reproducible scripts for the experiments are available at: https://github.com/Stefan-Maximilian-Dietrich/reliable-pls.

[3] The specific choice of model and distribution assumption here is more for the purpose of demonstration. The following can also be applied to other situations. In case of non-normal priors, alternative approximation strategies need to be considered.

The credal prior, which the Gamma-Maximin criterion is based on in our case, is a set of normal distributions that vary only in their location parameter, precisely

$$\Pi = \left\{ \pi_\mu : \pi_\mu = \mathcal{N}(\mu, \Sigma) \wedge \mu \in [\underline{\mu_0}, \overline{\mu_0}] \times \ldots \times [\underline{\mu_p}, \overline{\mu_p}] \right\}. \tag{3}$$

To calculate the marginal likelihood, it is not possible to resort to an analytical method. In this paper, we aim to adapt a numerical method that allows us to approximate the marginal likelihood with the help of the Laplace approximation. This is possible because the function for which the integral is to be computed is twice continuously differentiable and most of its mass is around its maximum [4,18]. We define $-h(\theta) := \log(\mathcal{L}(\theta) f_{\pi_\mu}(\theta))$ and $V(\theta)$ as the Hessian matrix of $h(\theta)$. Let $\tilde{\theta}$ be the argument that minimizes $h(\theta)$. This minimum is unique and global because $h(\theta)$ is convex. Now the Laplace approximation of the marginal likelihood[4] is:

$$m(\pi_\mu) \approx (2\pi)^{\frac{2}{d}} |V(\tilde{\theta})|^{-\frac{1}{2}} \mathcal{L}(\tilde{\theta}) f_{\pi_\mu}(\tilde{\theta}), \tag{4}$$

see [4,18]. To calculate $\tilde{\theta}$ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is used [5]. The BFGS method belongs to the quasi-Newton methods and is a numerical technique used to solve nonlinear optimization problems. This method appears particularly suitable for finding optima in our case, as our functions are twice differentiable and it saves computing resources by not computing the Hessian matrix in each step. To execute the soft revision we now want only the prior π_μ in our credal set that fulfills

$$m(\pi_\mu) - \alpha \max_{\pi_\mu \in \Pi} m(\pi_\mu) \geq 0 \tag{5}$$

We now want to construct an approximation of the pseudo posterior predictive (PPP), which depends only on the given data and on the prior π_μ. As described in [16], the PPP can be written as:

$$\frac{1}{p_{\pi_\mu}(\mathcal{D})} \int_\Theta p(\mathcal{D} \cup (x_i, \tilde{y}_i)|\theta) p(\mathcal{D}|\theta) f_{\pi_\mu}(\theta) d\theta \tag{6}$$

Here, we can approximate the marginal likelihood $p_{\pi_\mu}(\mathcal{D})$ again using the method described above because the difference is in $\mathcal{O}(n^{-1})$ [16]. The remaining integral is now similar to the marginal likelihood and can be approximated by Laplace because it is also twice continuously differentiable and most of its mass is around its mode. The approach here is analogous to the one used in approximating the marginal likelihood. Let now $q(\pi_\mu) \approx \int_\Theta p(\mathcal{D} \cup (x_i, \tilde{y}_i)|\theta) p(\mathcal{D}|\theta) f_{\pi_\mu}(\theta) d\theta$. The PPP dependent on the prior π_μ can now be approximated by the approximations of the marginal likelihood and the remaining integral:

$$p_{\pi_\mu}(\mathcal{D} \cup (x_i, \tilde{y}_i)|\mathcal{D}) \approx \frac{q(\pi_\mu)}{m(\pi_\mu)} \tag{7}$$

Now we have both a PPP for which it holds $\mathbb{E}_{\pi_\mu}(u(a, \theta)) = p_{\pi_\mu}(\mathcal{D} \cup (x_i, \tilde{y}_i)|\mathcal{D})$ dependent on a prior or its parameters and a soft revision (α-cut) criteria that

[4] d is the number of dimensions. The π without a subscript is the number π.

sorts out all priors that would overly restrict the influence of the likelihood. We now want to treat this as an optimization problem:

$$\begin{aligned} &\inf_{\pi_\mu \in \Pi} \mathbb{E}_{\pi_\mu}(u(a, \theta)) \\ &\text{s.t. } m(\pi_\mu) - \alpha \max_{\pi_\mu \in \Pi} m(\pi_\mu) \geq 0, \end{aligned} \tag{8}$$

which yields the Gamma-Maximin criterion with soft revision, see Eq. 2.

The optimization is implemented using the COBYLA (Constrained Optimization by Linear Approximations) algorithm [13]. The basic idea is forming linear approximations to the objective function via interpolation at the vertices of a simplex. This algorithm iteratively proposes new vertices for this simplex, similar to the Nelder-Mead algorithm. The COBYLA algorithm does not rely on gradients and handles inequality constraints well, which is useful for optimization of Eq. (8). As termination criterion, we use the fractional tolerance.

4 Experiments

We will now conduct various experiments. For this purpose, we estimate a model in each iteration, which we then evaluate using a test dataset. Since in each round the respective criteria need to be estimated for all pseudo-labeled data, the complexity lies in $\mathcal{O}((m-n)^2)$ with $m-n$ the number of unlabeled data. We benchmark semi-supervised logistic regression with the Gamma-Maximin method under soft revision (α-cut) (●) against the following methods: 1.) A standard supervised (●) baseline where only the labeled data is considered and no self-training takes place. 2.) three standard pseudo-label selection criteria: Probability Score (●), Predictive Variance (●) and Likelihood (max-max) (●). 3.) Four (robustified) Bayesian methods that depend on the pseudo posterior predictive: PPP (Bayes-optimal) (●), PPP multi-label (●), PPP multi-model (●) and PPP weighted multi-label (●).[5] Experiments are run on simulated binomially distributed data as well as on three data sets (Banknote, MTcars, Mushrooms) for binary classification from the UCI repository [8]. The binomially distributed data was simulated through a linear predictor consisting of normally distributed features. Furthermore, the effect of varying α is tested. Further results can be viewed in the repository.[6]

5 Discussion

At first glance, it is apparent that if any of the nine tested methods achieves an improvement over supervised learning, then Gamma-Maximin method with

[5] They incorporate uncertainty regarding model choice (multi-model and weighted multi-model (●)) and accumulation of errors (multi-label (●)). For a thourough introduction, please refer to [17].

[6] https://github.com/Stefan-Maximilian-Dietrich/reliable-pls.

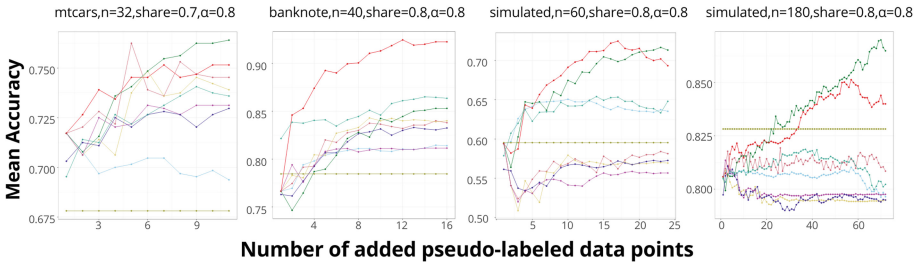


Fig. 1. Experiments on real-world data (left) and simulated data (right).

soft revision also achieves an improvement. In the conducted tests, it is evident that the Gamma-Maximin method with soft revision consistently ranks among the best out of all nine tested methods. This is particularly true for real-world data. In the case of simulated data, we see that besides the Gamma-Maximin method with soft revision, only the Multi-Model method consistently performs better than the Supervised Learning baseline. Upon closer comparison, two effects emerge: 1) The lower the total data volume, the better the performance of the Gamma-Maximin method with soft revision. 2) The further away the alternative models are to reality, the better the Gamma-Maximin method with soft revision performs relative to the Multi-Model method (in the example shown here, the alternative models are intentionally chosen to be very favorable to Multi-Model). Additionally, it was observed in the experiments that the Gamma-Maximin method with soft revision copes particularly well with a low proportion of labeled data compared to other methods. In the conducted experiments, no clear trend can be observed regarding which α leads to particularly good performance.

References

1. Aminian, G., Abroshan, M., Khalili, M., Toni, L., Rodrigues, M.: An information-theoretical approach to semi-supervised learning under covariate-shift. In: International Conference on Artificial Intelligence and Statistics, PMLR, pp. 7433–7449. (2022)
2. Augustin, T., Schollmeyer, G.: Comment: on focusing, soft and strong revision of choquet capacities and their role in statistics. Stat. Sci. **36**(2), 205–209 (2021)
3. Augustin, T., Coolen, F., De Cooman, G., Troffaes, M.: Introduction to Imprecise Probabilities. Wiley (2014)
4. Bleistein, N., Handelsman, R.A.: Asymptotic Expansions of Integrals. Ardent Media (1975)
5. Broyden, C.: The convergence of a class of double-rank minimization algorithms 1. General Considerations. IMA J. Appl. Math. **6**(1), 76–90 (1970)
6. Cattaneo, M.: A continuous updating rule for imprecise probabilities. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part III, pp. 426–435. Springer (2014)
7. Chapelle, O., Schölkopf, B., Zien, A.: Semi-Supervised Learning. MIT Press (2006)

8. Dua, D., Graff, C.: UCI Machine Learning Repository (2017). http://archive.ics.uci.edu/ml
9. Jansen, C., Augustin, T., Schollmeyer, G.: Decision theory meets linear optimization beyond computation. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2017). Lecture Notes in Computer Science, vol. 10369, pp. 329–339. Springer (2017)
10. Levi, I.: On indeterminate probabilities. J. Philos. **71**, 391–418 (1974)
11. Lienen, J., Hüllermeier, E.: Credal self-supervised learning. In: Neural Information Processing Systems, vol. 34, pp. 14370–14382. (2021)
12. Noubiap, R., Seidel, W.: An algorithm for calculating γ-minimax decision rules under generalized moment conditions. Ann. Stat. **29**, 1094–1116 (2001)
13. Powell, M.: A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Advances in Optimization and Numerical Analysis: Mathematics and Its Applications, pp. 51–67. Springer Netherlands (1994)
14. Rizve, M., Duarte, K., Rawat, Y., Shah, M.: In defense of pseudo-labeling: an uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: International Conference on Learning Representations (2021)
15. Rodemann, J.: Pseudo label selection is a decision problem. In: Proceedings of the 46th German Conference on Artificial Intelligence. Springer (2023)
16. Rodemann, J., Goschenhofer, J., Dorigatti, E., Nagler, T., Augustin, T.: Approximately Bayes-optimal pseudo-label selection. In: Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI). PMLR, vol. 216, pp. 1762–1773. (2023a)
17. Rodemann, J., Jansen, C., Schollmeyer, G., Augustin, T.: In all likelihoods: robust selection of pseudo-labeled data. In: Proceedings of the 13th International Symposium on Imprecise Probability: Theories and Applications. PMLR, vol. 215, pp. 412–425. (2023b)
18. Ruli, E., Sartori, N., Ventura, L.: Improved laplace approximation for marginal likelihoods. Electron. J. Stat. **10**(2), 3986–4009 (2016)
19. Triguero, I., García, S., Herrera, F.: Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowl. Inf. Syst. **42**, 245–284 (2013)
20. Utkin, L., Augustin, T.: Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In: Proceedings of the Forth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '05), pp. 349–358. SIPTA (2005)
21. Vandewalle, V., Biernacki, C., Celeux, G., Govaert, G.: A predictive deviance criterion for selecting a generative model in semi-supervised classification. Comput. Stat. Data Anal. **64**, 220–236 (2013)
22. Zhu, X.: Semi-supervised learning literature survey. Tech. Rep. 1530. Computer Science, University of Wisconsin-Madison, pp. 2. (2008)

Dissimilarity-Based Clustering with Soft Proximity Constraints

Fabrizio Durante[1] and Roberta Pappadà[2(✉)]

[1] Università del Salento, Dipartimento di Matematica e Fisica "Ennio De Giorgi", Lecce, Italy
fabrizio.durante@unisalento.it

[2] Università degli Studi di Trieste, Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche "B. de Finetti", Trieste, Italy
rpappada@units.it

Abstract. We provide a general framework for the definition of dissimilarity measures to cluster dependent random variables. A copula-based pairwise dissimilarity measure that takes into account soft proximity constraints is introduced and its properties are illustrated. Furthermore, some recent issues that are particularly of interest for the analysis of geo-referenced data are discussed.

Keywords: Cluster analysis · Copulas · Semi-supervised learning

1 Introduction

A clustering algorithm aims at identifying dependencies and association within a set $\mathcal{X}$ of $n \geq 3$ random variables in order to create a partition of $\mathcal{X}$ so that more similar variables belong to the same subset. To group variables, we usually introduce a way to measure the distances relative to the elements of $\mathcal{X}$ to decide which elements belong to a group. This is also called similarity, although on many occasions it is more convenient to consider a dissimilarity measurement.

Clustering methods are usually part of the unsupervised learning procedure [11], and their goal is to discover some underlying structure in the set of variables. However, it could also be the case that the set of random variables is observed with some background information expressed in terms of pairwise positive/negative relationships with an associated degree. Such latter information can provide some constraints to the whole clustering procedure. Thus, in such a case, we refer more properly to semi-supervised learning [2].

When dealing with time series clustering, for instance, several studies have recognized the importance of taking spatial information into account for geographically referenced samples. In such cases, it may be of interest to obtain clusters that show geographical proximity to enhance their interpretability [9,12].

The goal of this paper is to review some aspects of dissimilarity-based clustering of variables in a copula-based framework [10]. These methods take into

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 118–125, 2024.
https://doi.org/10.1007/978-3-031-65993-5_14

account both a dissimilarity measure that is rank-invariant and some constraints that can be interpreted in terms of proximity among the variables (as measured via spatial information and/or a set of covariates). As a specific aspect of our interest, we do not impose that the clustering output *must* satisfy the proximity constraints, but it *may* separate objects that are very similar into different clusters if they are spatially apart. This latter aspect can be referred to as soft proximity constraints [3].

In the following, first, we provide a general framework for dissimilarity-based clustering and we introduce a pairwise dissimilarity measure that takes into account soft proximity constraints (see Sect. 2). Then we discuss several examples in a copula-based setting with a particular focus on geo-referenced data (see Sect. 3). Finally, we present some final remarks about some novel approaches to the problem (see Sect. 4).

2 Definition of a Constrained Dissimilarity Measure

We denote by $L^0(\Omega, \mathcal{F}, \mathbb{P})$ the space of all continuous random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider a set $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ of $n \geq 3$ elements of $L^0(\Omega, \mathcal{F}, \mathbb{P})$.

Usually, we have at disposal a sample $\{(x_{t1}, \ldots, x_{tn})\}_{t=1,\ldots,T}$ from the probability law of $(X_1, \ldots, X_n)$ that is formed of iid observations. In fact, in the case the sample constitutes a (multivariate) time series, we can fit a standard copula-based model to the multivariate time series and extract the dependence information from the obtained residuals from the fitted model [14].

The following definition is derived from [10].

Definition 1. *A copula-based (pairwise) dissimilarity measure is a mapping*

$$\delta \colon L^0(\Omega, \mathcal{F}, \mathbb{P}) \times L^0(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}_+ := [0, +\infty[\tag{1}$$

that is invariant under increasing transformations of the input variables (X_1, X_2) *and, as such, depends only on their (unique) copula of* $C_{12} := C_{X_1,X_2}$*. Moreover, it satisfies the following additional properties:*

(D1) $\delta(X_1, X_2) = 0$ *if* X_1 *and* X_2 *are comonotonic;*
(D2) $\delta(X_1, X_2) = \delta(X_2, X_1)$ *for any random pair* (X_1, X_2)*;*
(D3) δ *is continuous with respect to the weak convergence of the joint probability law.*

Typical examples of dissimilarity measures are provided by the mapping

$$\delta(X_1, X_2) = 1 - \kappa(X_1, X_2),$$

where κ is a measure of concordance (like Kendall's τ or Spearman's ρ).

To take into account the additional information about the elements of $\mathcal{X}$, we assume that each variable X_i is equipped with a vector of p features $\mathbf{s}_i \in \mathbb{R}^p$,

$i = 1, \ldots, n$. In spatial statistics, these vectors may represent the geographic coordinates where the variables have been observed. To each pair of feature vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$, we can hence associate a proximity value $w_{ij} \geq 0$ given by the Euclidean distance (or any similar distance) between $\mathbf{s}_i$ and $\mathbf{s}_j$.

The dissimilarity measure in Definition 1 can be generalized to a mapping that includes the proximity value between the variables. Formally, we define a constrained (pairwise) dissimilarity measure as follows.

Definition 2. *A constrained copula-based (pairwise) dissimilarity measure is defined as a mapping*

$$\tilde{\delta}\colon L^0(\Omega, \mathcal{F}, \mathbb{P}) \times L^0(\Omega, \mathcal{F}, \mathbb{P}) \times S \to \mathbb{R}_+ \tag{2}$$

such that, for any $w \in S \subseteq \mathbb{R}_+$, $\tilde{\delta}(\cdot, \cdot, w)$ is a dissimilarity measure as in Definition 1 and, moreover,

(D4) $\tilde{\delta}(X_1, X_2, w) \leq \tilde{\delta}(X_1, X_2, w')$ if $w \leq w'$ for any fixed (X_1, X_2).

Thus, $\tilde{\delta}(X_1, X_2, w)$ can be interpreted as the dissimilarity value between X_1 and X_2 taking into account that their (spatial) proximity is equal to w. In view of property (D4), the constrained dissimilarity measure may up-weigh the small dissimilarities that may occur between sites that are far away or, equivalently, down-weigh the large dissimilarities that may occur between close neighbors.

Remark 1. The set S should be equipped with an order relation (here, for simplicity, it is a subset of the non-negative real line). For instance, one can also consider $S = \{0, 1\}$ so that $w = 0$ when two variables are contiguous, while $w = 1$, otherwise.

Since we are focusing on rank-invariant procedures, the constrained dissimilarity of Eq. (2) can be equivalently written as a function

$$\tilde{\delta}\colon \mathcal{C} \times S \to \mathbb{R}_+, \tag{3}$$

where $\mathcal{C}$ denotes the space of bivariate copulas [8]. A possible strategy to derive a function of type (3) has been suggested in [7] and is illustrated below.

Algorithm 1.
Let (X_i, X_j) be a random pair with copula C_{ij} such that $w_{ij} \in [0, 1]$ represents the proximity of X_i, X_j. To construct a dissimilarity measure of type (3), we proceed as follows:

1. Define a copula $C_{w_{ij}} \in \mathcal{C}$ that, roughly speaking, maps the proximity value w_{ij} into a proximity copula.
2. Mix the copulas C_{ij} and $C_{w_{ij}}$ by using a suitable operation φ in $\mathcal{C}$ (or in some of its subclasses). The resulting copula is denoted by $\widetilde{C}_{ij} = \varphi(C_{ij}, C_{w_{ij}})$.
3. For a dissimilarity measure δ as in Definition 1 define

$$\tilde{\delta}(C_{ij}, w_{ij}) = \delta(\varphi(C_{ij}, C_{w_{ij}})). \tag{4}$$

The following examples illustrate this approach.

Example 1. Let C_{ij} denote the copula associated with the i–th and the j–th variable and let w_{ij} denote their (normalized) distance. In [7] Algorithm 1 is specified in the following way:

1. The copula $C_{w_{ij}} = w_{ij}W + (1 - w_{ij})M$, where M and W are the comonotonicity and countermonotonicity copulas, respectively.
2. The copula $\widetilde{C}_{ij}$ is the convex combination of C_{ij} and $C_{w_{ij}}$ with a fixed weight $\gamma \in [0, 1]$.
3. The dissimilarity measure δ is defined as a distance between the copula M and the input copula C via an increasing and continuous real-valued function f and a suitable norm $\|\cdot\|$ in $\mathcal{C}$. Thus,

$$\tilde{\delta}(C_{ij}, w_{ij}) = f(\|\gamma(M - C_{ij}) + w_{ij}(1 - \gamma)(M - W)\|). \tag{5}$$

The function $f(\cdot)$ is such that $f(0) = 0$, hence the dissimilarity is zero for spatially adjacent pairs ($w \approx 0$) that are comonotone ($C_{ij} = M$). Moreover, for a fixed γ and copula C_{ij}, the dissimilarity increases as the spatial separation w_{ij} increases.

Example 2. In [1], a clustering method is presented with the aim of clustering time series of maxima. Each time series is associated with a vector of spatial features (i.e. geographical coordinates and elevation) describing the site at which the time series are collected. Let C_{ij} denote the copula associated with the i–th and the j–th time series and let w_{ij} denote their (normalized) distance.

In the context of maxima observations, it is natural to assume that C_{ij} belongs to the extreme-value family of copulas $\mathcal{C}_{EV}$. Thus, in order to develop a methodology that remains related to extreme-value copulas, Algorithm 1 has been modified in the following way:

1. Define the extreme-value copula $C_{w_{ij}}$ given by

$$C_{w_{ij}} = \Pi(u^{w_{ij}}, v^{w_{ij}})M(u^{1-w_{ij}}, v^{1-w_{ij}})$$

   where Π and M denote the independence and the comonotonicity copula, respectively.
2. Mix the copulas C_{ij} and $C_{w_{ij}}$ by using the operation on $\mathcal{C}_{EV}$ known as Khoudraji's device (see Eq. 4 in [1]) and given by

$$\varphi_\gamma(C_{EV}, C'_{EV})(u, v) = C_{EV}(u^{1-\gamma}, v^{1-\gamma})C'_{EV}(u^\gamma, v^\gamma),$$

   for a tuning parameter $\gamma \in [0, 1]$. The resulting copula is denoted by $\widetilde{C}_{ij}$.
3. Consider the dissimilarity measure $\delta(C) = -\ln(\lambda_U(C))$, where $\lambda_U(C)$ denotes the upper tail dependence coefficient associated with C (see [4]). Then it is proved that

$$\begin{aligned}\tilde{\delta}(C, w_{ij}) &= -\ln(\lambda_U(\varphi_\gamma(C_{ij}, C_{w_{ij}})) \\ &= -\ln((1 - \gamma)\lambda_U(C_{ij}) + \gamma(1 - w_{ij})).\end{aligned}$$

Clearly, for $C_{ij} = C_w = M$, with M denoting the comonotonicity copula, we have $\tilde{\delta}(C_{ij}, w) = 0$; moreover, the dissimilarity is increasingly monotonic in w.

3 A General Method to Obtain Constrained Dissimilarity Measures

For clustering with non-strict constrained procedures, a classical approach suggests running clustering algorithms on a modified dissimilarity matrix, which is a combination of the matrix computed from the variables and a matrix related to the geographical distances [13]. In this perspective, a constrained dissimilarity measure of type (2) can be constructed via the formula

$$\tilde{\delta}(X_i, X_j, w_{ij}) = f_\alpha(w_{ij}, \delta(X_i, X_j)),$$

where δ is a dissimilarity measure of Definition 1, and f_α is an increasing and continuous real-valued function, which depends on an additional parameter α that tunes the relative influence of the proximity value w on the whole calculation. This approach is illustrated via the following examples.

Example 3. Several spatial weighting functions have been explored in the literature to incorporate geographical constraints in the definition of a suitable dissimilarity matrix. As discussed, for instance, in [13], the original dissimilarity may be modified by introducing a nonlinear function of the separating distance between any pair of sites to increase the similarity between points that are spatially close, without affecting the similarity between distant points. One way to do this is to consider a dissimilarity of the form

$$\tilde{\delta}(X_i, X_j, w_{ij}) = g_\alpha(w_{ij})\delta(X_i, X_j),$$

where $g_\alpha(w) = 1 - \exp(-w/\alpha)$, $w_{ij} \geq 0$ is the distance between the two sites, and $\alpha > 0$ is an arbitrary weight. It follows that the pair (X_i, X_j) has dissimilarity equal to zero either when they are comonotonic, or when they are spatially contiguous, i.e. when $w_{ij} = 0$. Moreover, for every $w_{ij} \geq 0$, it holds $\tilde{\delta}(X_i, X_j, w_{ij}) \leq \delta(X_i, X_j)$. Thus short distances result in small dissimilarities, even for the random variables that exhibit large dissimilarity.

Example 4. When dealing with spatio-temporal data, a simple way to build a constrained dissimilarity is to consider a convex combination of the temporal dissimilarity matrix and a spatial matrix, as done for instance in [5]. In the copula space, the constrained copula-based dissimilarity measure can be derived similarly via the formula:

$$\tilde{\delta}(X_i, X_j, w_{ij}) = \alpha w_{ij} + (1 - \alpha)\delta(X_i, X_j),$$

where δ is a dissimilarity of Definition 1, but scaled to a value between 0 and 1, and w_{ij} is the normalized proximity. The mixing parameter $\alpha \in [0,1]$ represents the weight associated with the spatial closeness and the case $\alpha = 0$ corresponds to the absence of spatial influence. Notice that, for a fixed α and for every $w_{ij} \geq 0$, $\tilde{\delta}(X_i, X_j, w_{ij}) \geq \delta(X_i, X_j)$, with $\tilde{\delta}(X_i, X_j, 0) = \delta(X_i, X_j)$. Hence, the dissimilarity penalizes the aggregation of pairs having a large spatial separation.

Example 5. The constrained dissimilarity measure proposed by [6] is defined as a function of the parameter of the copula adopted to describe the degree of cross-sectional dependence among the involved times series. Specifically, the dissimilarity between any pair of random variables is of the form

$$\tilde{\delta}(X_i, X_j, w_{ij}) = \exp(w_{ij})\delta(X_i, X_j) = \exp(w_{ij})\sqrt{2(1 - \theta_{C_{ij}})},$$

where $\theta_{C_{ij}} \in [-1, 1[$ is the dependence parameter of the AMH copula fitted to (X_i, X_j), and $w_{ij} \in [0,1]$ is the normalized geographical distance. Hence, the dissimilarity degree tends to vanish as approaching the comonotonic (positive) case. Moreover, the dissimilarity accounts for a penalization term towards objects that are further apart. Clearly, $\tilde{\delta}(X_i, X_j, w_{ij}) \geq \delta(X_1, X_2)$, and $\tilde{\delta}(X_i, X_j, 0) = \delta(X_i, X_j)$.

Example 6. A further example can be found in [15], where the dissimilarity measure takes into account the co-occurrence of extreme events as well as the spatial contiguity between the sites where data are collected. In such a case

$$\tilde{\delta}(X_i, X_j, w_{ij}) = -\ln(\lambda_U(C_{ij})) + \alpha w_{ij}$$

where w_{ij} is a contiguity coefficient assuming value 0 when the pair (X_i, X_j) comes from neighbouring areas, and 1 otherwise; moreover, the parameter α adjusts the impact of the contiguity coefficient in the dissimilarity between two time series (see [15, Section 4] for further details). The dissimilarity defined above satisfies $\tilde{\delta}(X_i, X_j, w_{ij}) \geq \delta(X_i, X_j)$, for all $w_{ij} \geq 0$, and $\tilde{\delta}(X_i, X_j, 0) = \delta(X_i, X_j) = -\ln(\lambda_U(C_{ij}))$.

In most of the previous examples, a crucial role is played by a tuning parameter $\alpha \in [0,1]$. As clarified in [3], according to the weight α given to the geographical dissimilarities, the cluster output will have more or less spatially contiguous clusters. However, this approach raises the problem of defining weight in an objective manner (see also [9]).

4 Final Remarks

We have provided a review of the theoretical framework for the construction of dissimilarity measures in clustering algorithms used to group variables. In

particular, we presented a general approach to build dissimilarity functions for continuous random vectors that are based on the use of copulas, and we have discussed their properties. We focused on those dissimilarities that allow us to incorporate the spatial information available for the data, by introducing the notion of constrained copula-based dissimilarity measure. A major issue arising from the literature is the choice of the tuning parameter ruling the importance of spatial proximity. This aspect deserves further investigation, since it may affect the quality of the final clustering.

Acknowledgements. FD has been supported by MUR-PRIN 2022 PNRR, Project "Stochastic Modeling of Compound Events" (No. P2022KZJTZ) funded by European Union—Next Generation EU. The work of FD has been carried out with partial financial support from ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by EU—Next Generation EU (CUP F83C22000740001). RP has been supported by MUR-PRIN 2022, Project "Modelling Non-standard data and Extremes in Multivariate Environmental Time series" (No. 20223CEZSR) funded by European Union—Next Generation EU.

References

1. Benevento, A., Durante, F., Pappadà, R.: An approach to cluster time series extremes with spatial constraints. In: Chelli, F., Ciommi, M., Ingrassia, S., Mariani, F., Recchioni, M. (eds.) Book of Short Papers SEAS IN 2023, pp. 679–684. Pearson (2023)
2. Cai, J., Hao, J., Yang, H., Zhao, X., Yang, Y.: A review on semi-supervised clustering. Inf. Sci. **632**, 164–200 (2023)
3. Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J.: ClustGeo: an R package for hierarchical clustering with spatial constraints. Comput. Stat. **33**(4), 1799–1822 (2018)
4. De Luca, G., Zuccolotto, P.: A tail dependence-based dissimilarity measure for financial time series clustering. Adv. Data Anal. Classif. **5**(4), 323–340 (2011)
5. Deb, S., Karmakar, S.: A novel spatio-temporal clustering algorithm with applications on covid-19 data from the united states. Comput. Statist. Data Anal. **188**, 107810 (2023)
6. Di Lascio, F.M.L., Menapace, A., Pappadà, R.: A spatially-weighted AMH copulabased dissimilarity measure for clustering variables: an application to urban thermal efficiency. Environmetrics **35**(1), e2828 (2024)
7. Disegna, M., D'Urso, P., Durante, F.: Copula-based fuzzy clustering of spatial time series. Spat. Stat. **21**(part A), 209–225 (2017)
8. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton, FL (2016)
9. Fouedjio, F.: Clustering of multivariate geostatistical data. WIREs Comput. Stat. **12**(5), e1510 (2020)
10. Fuchs, S., Di Lascio, F.M.L., Durante, F.: Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Comput. Statist. Data Anal. 107201 (2021)

11. Hennig, C., Meila, M., Murtagh, F., Rocci, R.: In: Handbook of Cluster Analysis. Chapman Hall/CRC Handbook of Modern Statistical Methods. Boca Raton, FL, CRC Press (2016)
12. Kopczewska, K.: Spatial machine learning: new opportunities for regional science. Ann. Reg. Sci. **68**(11), 713–755 (2022)
13. Oliver, M.A., Webster, R.: A geostatistical basis for spatial weighting in multivariate classification. Math. Geol. **21**(1), 15–35 (1989)
14. Patton, A.: A review of copula models for economic time series. J. Multivariate Anal. **110**, 4–18 (2012)
15. Zuccolotto, P., De Luca, G., Metulini, R., Carpita, M., et al.: Modeling and clustering of traffic flows time series in a flood prone area. In: Proceedings of the Statistics and Data Science Conference, pp. 113–118. EGEA (2023)

Copula-Based Fuzzy Clustering of Count Data with Total Variation Distance

Pierpaolo D'Urso[1], Livia De Giovanni[2], Lorenzo Federico[2(✉)], and Vincenzina Vitale[1]

[1] Department of Social Science and Economics, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Lazio, Italy
{pierpaolo.durso,vincenzina.vitale}@uniroma1.it

[2] Department of Political Sciences and Data Lab, Luiss University, Viale Romania, 32, 00197 Rome, Lazio, Italy
{ldegiovanni,lfederico}@luiss.it

Abstract. This paper presents a novel fuzzy clustering technique designed specifically for count data, referred to as the Fuzzy C-medoids algorithm based on Total Variation Distance. We evaluate its performance against a benchmark relying on Shannon divergence, commonly employed in scenarios involving discrete probability distributions, through simulation analysis. A comprehensive evaluation of the proposed approach's effectiveness is carried out, revealing promising results. The study's findings emphasize the potential of the proposed fuzzy method, particularly in scenarios where discrete probability distributions are involved.

Keywords: Total variation distance · Fuzzy C-medoids · Count data

1 Introduction

Clustering of count data involves grouping observations based on their count values, *i.e.* based on the similarity of count patterns, rather than continuous variables. It requires methods that can handle discrete and often non-negative data effectively. Traditional methods like K-Means or Hierarchical clustering using Euclidean metrics are insufficient, as distance metrics tailored to the characteristics of count data are required.

While there is extensive literature available on model-based clustering methods for count data, the same cannot be said for model-free clustering approaches, also in the fuzzy context. The latter class of methods is suitable for datasets where points may have ambiguous cluster assignments or belong to multiple clusters simultaneously. In this paper, we consider the well-known extension of Fuzzy C-means (FCM, [1,2,4]), the so-called Fuzzy C-Medoids clustering method (FCMd [7,8]).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 126–133, 2024.
https://doi.org/10.1007/978-3-031-65993-5_15

Employing the Partitioning Around Medoids algorithm, it distinguishes itself from traditional Fuzzy C-Means (FCM) by identifying actual observed objects from the dataset, known as medoids, as prototypes for each cluster. This distinction is significant as medoids, being robust representatives of clusters, offer greater stability and reliability in the presence of noise and outliers, thereby making FCMd particularly suitable for scenarios where outlier impact on clustering outcomes is substantial. In this paper, we propose a suitable extension of FCMd using the Total Variation distance, a metric that satisfies several desirable properties and is well-suited for comparing dissimilarities between distributions, including discrete ones. Thanks to its connection with copula models, our proposed extension outperforms Shannon divergence [9], particularly in cases where clusters are generated based on non-linear dependence structures, as demonstrated through simulations.

The outline of the paper is structured as follows: In Sect. 2, we introduce the Total Variation distance. Subsequently, in Sect. 3, we detail the proposed clustering method. Section 4 presents the results of our simulations. Finally, the last section discusses potential future extensions.

2 Total Variation Distance

Let us denote two probability measures over the finite set Ω as μ and η. The total variation distance between these two probability measures is defined as

$$\|\mu - \eta\|_{\mathrm{TVD}} = \sup_{A \subseteq \Omega} |\mu(A) - \eta(A)| \tag{1}$$

where $\mu(A)$ and $\eta(A)$ are the probability of the event A by the probability measures μ and η respectively. In simpler terms, it represents the supremum of the absolute differences in probabilities assigned by the two measures to all possible events A.

For discrete distribution, it can be reformulated as the the normalized L1-distance between the two probability measures, that is,

$$\|\mu - \eta\|_{\mathrm{TVD}} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)| \tag{2}$$

where $\mu(\omega)$ and $\eta(\omega)$ represent the probabilities assigned to the element ω in the sets Ω by the probability measures μ and η respectively. This is a metric assuming values between 0 and 1, with 0 denoting identical distributions and 1 signifying complete non-overlap.

A key property of the Total Variation Distance is that, for all bivariate random variables (X, Y) with marginal distributions $X \sim \mu$ and $Y \sim \eta$, we have $\mathbb{P}(X \neq Y) \geq \|\mu - \eta\|_{\mathrm{TVD}}$ and there exists a copula C such that if (X, Y) follows the distribution $C(\mu, \eta)$, then $\mathbb{P}(X \neq Y) = \|\mu - \eta\|_{\mathrm{TVD}}$ (cfr. [6]).

3 The Proposed Clustering Method

Let $\mathbf{X} = \{x_{ij}, i = 1, \dots I, j = 1 \dots k\}$ a matrix of count data, of dimension $I \times k$, *i.e.* each $\mathbf{x}_i$ can be seen as a realization of a multinomial process, and let $\mathbf{F} = \{f_{ij}, i = 1, \dots I, j = 1 \dots k\}$ the corresponding matrix of row profiles, such that $\mathbf{f}_i = (f_{i1}, f_{i2}, \dots f_{ik})$.

The proposed Fuzzy C-Medoids clustering method, labeled $FCMd_{TVD}$, applied to the $\mathbf{F}$ matrix, uses the distance (2) to identify a subset of C prototypes, called medoids, along with the matrix of fuzzy coefficients denoted as $\mathbf{U}_{n \times C}$. This is achieved through the minimization of the following objective function:

$$\begin{cases} min : \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^{m} \|\mathbf{f}_i - \widetilde{\mathbf{f}}_c\|_{\text{TVD}}^2 \\ \qquad = \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^{m} [\frac{1}{2} \sum_{j=1}^{k} |f_{ij} - \widetilde{f}_{cj}|]^2 \\ \quad s.t. \sum_{c=1}^{C} u_{ic} = 1, \; u_{ic} \geq 0. \end{cases} \tag{3}$$

$u_{ic} \in [0, 1]$ represents the membership degree of the *i-th* unit in the *c-th* cluster, while $\widetilde{\mathbf{f}}_c$ is the medoid for cluster c. The "fuzziness parameter", denoted as "m" is a critical factor in controlling the level of fuzziness during the clustering process. Higher values of m result in fuzzier memberships, and it is usually set above 1 to guarantee algorithm convergence (for additional details on the role of m, refer to [5]).

The solution for each u_{ic}, for $i = 1, \dots I$ and $c = 1 \dots C$ is:

$$u_{i,c} = \frac{1}{\sum_{c'=1}^{C} \left[\frac{(\sum_{j=1}^{k} |f_{ij} - \widetilde{f}_{cj}|)^2}{(\sum_{j=1}^{k} |f_{ij} - \widetilde{f}_{c'j}|)^2} \right]^{\frac{1}{m-1}}} \tag{4}$$

4 Simulation Study

A simulation plan has been implemented to evaluate the effectiveness of the proposed fuzzy approach. For benchmarking purposes, we utilized the C-medoids fuzzy clustering method based on the Shannon divergence, henceforth named as $FCMd_{SH}$, since it is commonly applied in scenarios involving discrete probability distributions.

Two distinct scenarios have been outlined, each stemming from separate data generation processes. The first scenario involves simulating count data using the multinomial distribution, while the second scenario uses copulas for data generation.

First simulation We simulated a matrix $\mathbf{X}$ of dimension $I = 47$ and $k = 4$ considering the following structure of groups.

- Cluster 1: $\mathbf{x}_i$, for $i = 1, \dots 15$, are generated from a Multinomial distribution with vector of probabilities $\mathbf{p} \equiv (0.10, 0.10, 0.10, 0.70)$ and n randomly generated from a uniform distribution on the interval $[40, 200]$

- Cluster 2: $\mathbf{x}_i$, for $i = 1, \ldots 15$, are generated are generated from a Multinomial distribution with vector of probabilities $\mathbf{p} \equiv (0.5, 0.3, 0.1, 0.1)$ and n randomly generated from a Poisson distribution with mean 80
- Cluster 3: $\mathbf{x}_i$, for $i = 1, \ldots 15$, are generated from a Multinomial distribution with vector of probabilities $\mathbf{p} \equiv (0.1, 0.15, 0.35, 0.35)$ and n generated from a geometric distribution with parameter 0.007 for $i = 1, \ldots 10$, from a Poisson distribution with mean 80, for $i = 11, \ldots 15$.

Two further units are added with generating count profiles $\mathbf{x}$ obtained as $x_{i1} \sim U(180, 200)$, $x_{i2} \sim U(1, 10)$, $x_{i3} \sim U(1, 20)$, $x_{i4} \sim U(1, 10)$, for $i = 1, 2$. These two units exhibit highly skewed profiles compared to the majority of the data, resembling outlier units. This evaluation aims to assess robustness in the presence of outliers. *Second simulation* We simulated a matrix $\mathbf{X}$ of dimension $I = 52$ and $k = 4$ considering the following structure of groups.

- Cluster 1: we generated a sample of 20 observations from a Gumbel copula with parameter $\theta = 4$ and margins given by Poisson distributions with parameters $\lambda_1 = 0.2 * r$, $\lambda_2 = 0.5 * r$, $\lambda_3 = 0.1 * r$ and $\lambda_4 = 0.2 * r$ respectively, with $r \sim U(50, 100)$
- Cluster 2: we generated a sample of 15 observations from a Gaussian copula with correlation parameter $\rho = 0.8$ for each couple of variables and margins given by Poisson distributions with parameters $\lambda_1 = 0.10 * r$, $\lambda_2 = 0.2 * r$, $\lambda_3 = 0.5 * r$ and $\lambda_4 = 0.2 * r$ respectively, with $r \sim U(50, 100)$
- Cluster 3: we generated a sample of 15 observation from a Frank copula with parameter $\theta = 6$ and margins given by Poisson distributions with parameters $\lambda_1 = 0.5 * n$, $\lambda_2 = 0.2 * r$, $\lambda_3 = 0.15 * r$ and $\lambda_4 = 0.15 * r$ respectively,with $r \sim U(50, 100)$

Two further units are added with generating count profiles $\mathbf{x}$ obtained as $x_{i1} \sim U(180, 200)$, $x_{i2} \sim U(1, 20)$, $x_{i3} \sim U(2, 10)$, $x_{i4} \sim U(2, 20)$ for $i = 1, 2$. Also, in this case, these two units exhibit highly skewed profiles compared to the majority of the data, resembling outlier units.

We present an example of simulated row profiles in Fig. 1 following the aforementioned scheme, based on three out of four dimensions.

An additional sub-scenario has been explored, wherein the two extra units are introduced, each with distinct generating count profiles, *i.e.* $x_{11} \sim U(180, 200)$, $x_{12} \sim U(1, 20)$, $x_{13} \sim U(2, 10)$, $x_{14} \sim U(2, 20)$ and $x_{21} \sim U(2, 20)$, $x_{22} \sim U(1, 10)$, $x_{23} \sim U(1, 20)$, $x_{24} \sim U(180, 200)$. Also, in this case, these two units exhibit highly skewed profiles, with the former displaying predominantly positive skewness and the latter exhibiting predominantly negative ones.

To validate the effectiveness of the proposed clustering method based on Total Variation distance, we executed the algorithm with $C = 3$ and $m = 1.3$, employing 100 random restarts on 100 simulated data sets. The same parameter values were used for the Shannon divergence-based clustering algorithm.

The Fuzzy Adjusted Rand Index ([3], FARI) was employed to compare the obtained fuzzy partition with the true one (excluding the last two units). The adjusted Rand index ranges from -1 to 1, where 1 indicates perfect agreement,

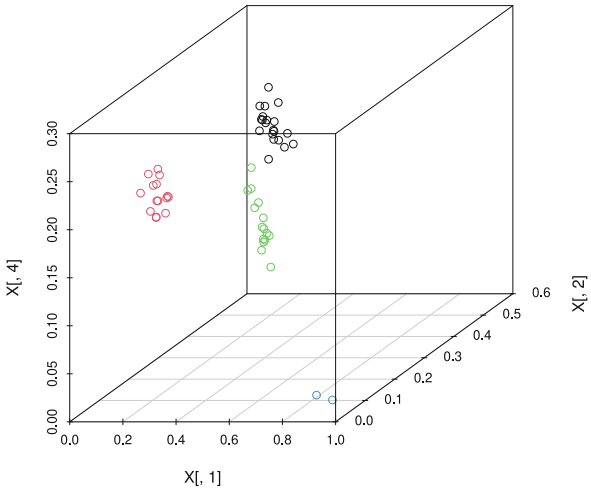


Fig. 1. An example of data based on the second simulation scheme with groups distinguished by distinct colors, with the two outliers highlighted in blue.

0 denotes agreement by chance, and negative values indicate disagreement worse than random. For each scenario, we computed key descriptive statistics for the Fuzzy Adjusted Rand Index across 100 simulations for every clustering method. These statistics are reported in Tables 1 and 2, respectively, encompassing the minimum value, first quartile, median, mean, third quartile, maximum value, and standard deviation. The corresponding boxplots are shown in Figs. 2 and 3.

Table 1. First simulation: minimum value, first quartile, median, mean, third quartile, maximum value and standard deviation of the Fuzzy adjusted Rand index over 100 simulations for both clustering methods

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	s.d.
$FCMd_{SH}$	0.463	1.000	1.000	0.988	1.000	1.000	0.069
$FCMd_{TVD}$,	0.932	0.997	0.999	0.995	1.000	1.000	0.013

The simulation findings clearly show that $FCMd_{TVD}$ outperforms the benchmark method in both scenarios, particularly when data are characterized by copula-based dependence structures. Its primary strength lies in its enhanced stability, as evidenced by consistently lower standard deviation values compared to those of the benchmark method.

In scenarios characterized by increased complexity, such as those featuring non-spherical clusters, employing Total Variation distance proves to be an effective solution. This observation suggests that Total Variation distance offers a robust approach for handling complex data structures and is particularly beneficial in situations where traditional methods might be inadequate. The use

Table 2. Second simulation: minimum value, first quartile, median, mean, third quartile, maximum value and standard deviation of the Fuzzy adjusted Rand index over 100 simulations for both clustering methods. The first two rows correspond to the first sub-scenario, while the last pair corresponds to the second sub-scenario.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	s.d.
$FCMd_{SH}$	0.394	0.492	1.000	0.809	1.000	1.000	0.258
$FCMd_{TVD}$	0.343	0.999	1.000	0.959	1.000	1.000	0.140
$FCMd_{SH}$	0.312	1.000	1.000	0.929	1.000	1.000	0.194
$FCMd_{TVD}$	0.465	1.000	1.000	0.983	1.000	1.000	0.090

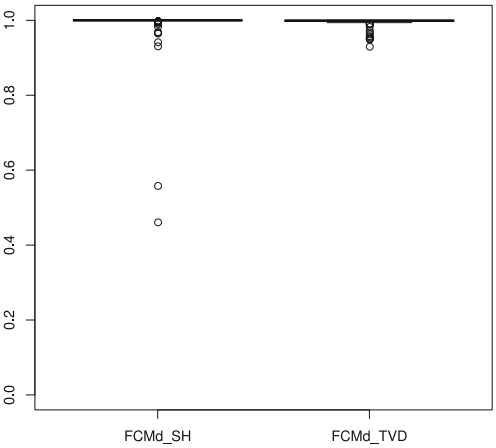


Fig. 2. First simulation: boxplot of the Fuzzy adjusted Rand index

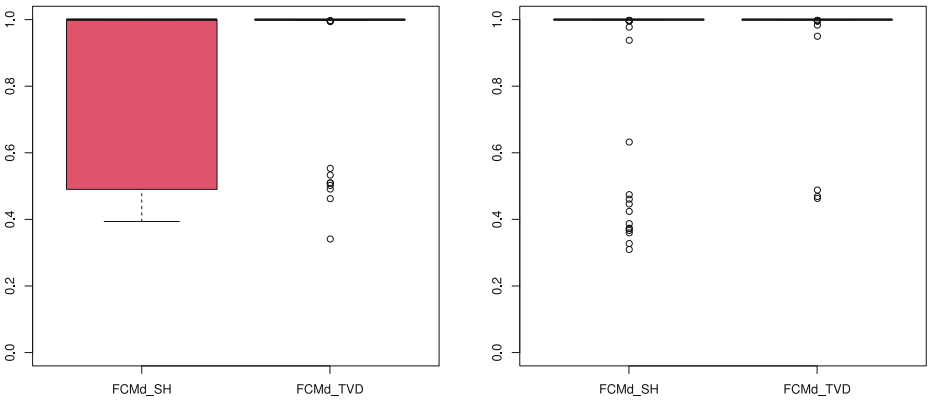


Fig. 3. Second simulation: boxplot of the Fuzzy adjusted Rand index for the first subscenario (on the left) and the second sub-scenario (on the right)

of a fuzzy approach presents a distinct advantage since the proposed method becomes well-suited for clustering complex data intrinsically characterized by inherent imprecision. This feature allows for the effective grouping of objects that exhibit varying levels of uncertainty or ambiguity, enhancing the adaptability and robustness of the approach in handling complex datasets.

The use of a fuzzy medoids-based approach enhances robustness to noise and outliers compared to traditional fuzzy clustering algorithms. Additionally, the ability to compute medoids using any dissimilarity measure renders the fuzzy medoids approach suitable for dataset where Euclidean distance metrics are not applicable. Furthermore, in some cases, the computational complexity of computing medoids can be lower than that of centroids, particularly for large datasets. This scalability advantage makes the fuzzy medoids approach more adaptable and applicable to high-dimensional datasets or large sample sizes.

5 Conclusions

The simulation analysis conducted in this study validates the effectiveness of the proposed fuzzy method in comparison to the benchmark fuzzy clustering method based on Shannon divergence. The proposed approach shows better performance in terms of both robustness and stability. The methodology inherits all the advantages associated with the chosen metric and medoid-based algorithm, which is known to be less sensitive to extreme values compared to the mean. The findings contribute to the ongoing advancement of fuzzy methodologies and offer practical implications for their implementation in real-world scenarios. Future research endeavors may explore extensions of the proposed method to very complex objects like histogram data.

References

1. Bezdek, J.C.: Numerical taxonomy with fuzzy sets. J. Math. Biol. **1**, 57–71 (1974)
2. Bezdek, J.C.: Objective function clustering. In: Pattern Recognition with Fuzzy Objective Function Algorithms, pp. 43–93. Advanced Applications in Pattern Recognition, Springer New York, NY (1981)
3. Campello, R.J.: A fuzzy extension of the rand index and other related indexes for clustering and classification assessment. Pattern Recognit. Lett. **28**(7), 833–841 (2007)
4. Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. J. Cybernetics **3**(3), 32–57 (1973)
5. D'Urso, P.: Fuzzy clustering. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (eds.) Handbook of Cluster Analysis, Chap. 24. Chapman and Hall/CRC (2015)
6. Kelbert, M.: Survey of distances between the most popular distributions. Analytics **2**(1), 225–245 (2023)
7. Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L.: Low-complexity fuzzy relational clustering algorithms for web mining. IEEE Trans. Fuzzy Syst. **9**(4), 595–607 (2001)

8. Krishnapuram, R., Joshi, A., Yi, L.: A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering. In: International Fuzzy Systems Conference (FUZZIEEE99), Seoul, Vol. 3, pp. 1281–1286. IEEE (1999)
9. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory **37**(1), 145–151 (1991)

Missing in Survival Analysis

Sam Efromovich and Lirit Fuksman(✉)

The University of Texas at Dallas, 800 W Campbell Rd., Richardson, TX 75080, USA
{efrom,lirit.fuksman}@utdallas.edu

Abstract. Typical data in survival analysis is right censored. Right censoring implies that some observations of the lifetimes of interest are not available. Additionally, survival data is often incomplete due to missing observations. Accordingly, it is important to develop optimal procedures that deal with missing censored data. Optimal estimation of bivariate density based on missing censored data is considered, and the theory is complemented by numerical simulations and an example.

Keywords: Adaptation · Density estimation · Integrated squared error · Missing data · Survival analysis

1 Introduction

The overall aim of the paper is to propose the methodology of nonparametric estimation of a bivariate density based on missing censored observations. We begin with a brief overview of known results related to the studied problem.

Let X be a continuous lifetime of interest supported on $[0, 1]$. We are interested in estimation of its α-fold differentiable density f^X. Based on a sample of size n from X, the density can be estimated with the optimal rate $n^{-2\alpha/(2\alpha+1)}$, and even sharp constant of the mean integrated squared error (MISE) convergence is known, see [3,13]. If we are interested in joint distribution of the pair (X, Y) of lifetimes supported on $[0, 1]^2$, and the joint density $f^{X,Y}$ is α-fold differentiable in each variable, then the optimal rate decreases to $n^{-\alpha/(\alpha+1)}$, and the decrease is called the "curse of dimensionality". At the same time, the same methodology is used for estimation of univariate and bivariate densities, see [4].

In survival analysis lifetimes of interest are rarely directly observable due to right censoring. For instance, in a cancer study with the lifetime of interest being the time from surgery to the cancer relapse, that time is right censored if the patient dies from other complications rather than from the treated cancer. Let C be the censoring lifetime, then the censored sample is from $(V, \Delta) := (\min(X, C), I(X \le C))$. Here $I(\cdot)$ is the indicator function and Δ is called the indicator of censoring. The famous Kaplan-Meier estimator (KME) of survival functions, based on the product-limit methodology, is the main tool for dealing with censored data. Further, the KME makes possible estimation of the density of X. Under a mild assumption, including independence between X and C, the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 134–141, 2024.
https://doi.org/10.1007/978-3-031-65993-5_16

censoring does not slow down the rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE convergence but affects the sharp constant, see [5]. Now consider a case when we are interested in a pair of lifetimes (X, Y). A typical example is the study of two successive courses of treatments. If right censoring occurs, then a sample of size n comes from a quartet $(V, W, \Delta, \Gamma) := (\min(X, C), \min(Y, D), I(X \le C), I(Y \le D))$. Here C and D are censoring times. Unfortunately, there is no straightforward extension of the univariate product-limit methodology to the bivariate case. Instead, a number of sophisticated and mathematically involved procedures, ranging from EM, hazard gradient, partial differential equations and copula to nonparametric MLE and solving an inhomogeneous Volterra equation via Peano series, have been proposed, see an insightful discussion in [1,2,10,11]. For the bivariate density estimation, it has been recently established that under a mild assumption the rate $n^{-\alpha/(\alpha+1)}$, known for direct observations, is preserved but the constant is increased, see [8].

Missingness is another familiar complication in survival analysis, see [12,14]. First, consider the univariate case. Let A be a Bernoulli random variable called the availability, and consider a sample from AX. Then it is shown in [6] that if $\mathbb{P}(A|X = x) = k > 0$, and this case is called missing completely at random (MCAR), then the optimal rate $n^{-2\alpha/(2\alpha+1)}$ is preserved. At the same time, if $\mathbb{P}(A|X = x) = k(x)$ then no consistent estimation is possible. Missingness in univariate censored data, when a sample is from (AV, Δ), is discussed in [7] where the methodology is suggested and the optimal univariate rate is attainable. No results about optimal estimation of bivariate density with missing lifetimes have been known so far.

Now we can formulate the main studied problem. We observe a sample of size n from the quintet $(AV, AW, \Delta, \Gamma, A)$. The problem is to estimate the underlying density $f^{X,Y}$ over a bounded rectangle, and without loss of generality we are considering the unit square $\mathcal{R} := [0, 1]^2$. The aim is to show that, under a mild assumption, the optimal rate $n^{-\alpha/(\alpha+1)}$, known for the case of direct observations from (X, Y), is still attainable for the missing censored data.

The context of the paper is as follows. The theory of rate optimal estimation is presented in Sect. 2. It is shown that the optimal rate can be preserved. The algorithm of bivariate density estimation for small samples can be found in Sect. 3, and simulations and a practical example are in Sect. 4.

Let us introduce the notation used throughout the paper. In what follows, $f^{X,Y}$ is the joint density of interest and $S^{C,D}(x, y) = P(C \ge x, D \ge y)$ is the joint survival function of censoring variables. Set $q_n = \lceil ln(n + 20) \rceil$, and let Q, a and b be positive constants. Expectation is denoted by $\mathbb{E}\{\cdot\}$ and indicator function is denoted by $I(\cdot)$.

2 Theory

We are considering estimation of the bivariate density $f^{X,Y}$ over the unit rectangle $\mathcal{R} := [0, 1]^2$ using a sample of size n from the quintet $(AV, AW, \Delta, \Gamma, A) := (A \min(X, C), A \min(Y, D), I(X \le C), I(Y \le D), A)$. Here (C, D) are censoring lifetimes and A is the Bernoulli random variable called the availability.

Assumption 1. The pair of continuous random variables (X, Y) is independent from a pair of continuous censoring random variables (C, D). The joint survival function $S^{C,D}(1,1) > 0$.

This assumption is necessary for consistent estimation.

The next assumption defines the smoothness of the underlying $f^{X,Y}$. Introduce the tensor-product basis $\{\varphi_{ji}(x,y) := \varphi_j(x)\varphi_i(y), i, j = 0, 1, \ldots\}$ on $\mathcal{R}$, $\varphi_j(x) = I(j = 0) + 2^{1/2}\cos(\pi j x)I(j \geq 1)$. Following [9], introduce the Sobolev class $\mathcal{S}(\alpha, Q) := \{f : \sum_{j,i=0}^{\infty}(1 + j^{2\alpha} + i^{2\alpha})\theta_{ji}^2 \leq Q < \infty,\ \theta_{ji} := \int_{\mathcal{R}} f(x,y)\varphi_{ji}(x,y)dydx\}$.

Assumption 2. Bivariate density $f^{X,Y}$ is from Sobolev class $\mathcal{S}(\alpha, Q)$.

The final assumption is about an underlying missing mechanism.

Assumption 3. The availability A is Bernoulli and $\mathbb{P}(A = 1|V, W, \Delta, \Gamma) = k > 0$ almost sure.

This missingness is the classical missing completely at random (MCAR), see [7,12].

Now we can explain the heuristic of rate-optimal estimation of $f^{X,Y}$. In what follows Bs are generic positive constants. Recall that we would like to match the upper bound $\sup_{f^{X,Y}\in\mathcal{S}(\alpha,Q)}\mathbb{E}\{\int_{\mathcal{R}}(\check{f}^{X,Y}(x,y) - f^{X,Y}(x,y))^2 dxdy\} \leq Bn^{-\alpha/(\alpha+1)}$ known for the case of direct observations of (X, Y). Using the tensor-product basis $\{\varphi_{ji}\}$ and the Fourier theorem we can write the estimand as the Fourier series $f^{X,Y}(x,y) = \sum_{i,j=0}^{\infty}\theta_{ji}\varphi_{ji}(x,y)$ where $\theta_{ji} := \int_{\mathcal{R}} f^{X,Y}(x,y)\varphi_{ji}(x,y)dydx$ are the Fourier coefficients. Let

$$\tilde{f}^{X,Y}(x,y) = \sum_{j,i=0}^{J}\tilde{\theta}_{ji}\varphi_{ji}(x,y). \tag{1}$$

Let J be the smallest integer larger than $n^{1/2(\alpha+1)}$. Suppose that the Fourier estimate $\tilde{\theta}_{ji}$ satisfy

$$\mathbb{E}\{(\tilde{\theta}_{ji} - \theta_{ji})^2\} \leq B_* n^{-1}, \quad B_* < \infty. \tag{2}$$

Using the Parseval identity, we write

$$\mathbb{E}\{\int_{\mathcal{R}}(\tilde{f}^{X,Y}(x,y) - f^{X,Y}(x,y))^2\} = \sum_{j,i=0}^{J}\mathbb{E}\{(\tilde{\theta}_{ji} - \theta_{ji})^2\} + \sum_{\max(j,i)>J}\theta_{ji}^2.$$

We can estimate the first sum using (2) and the second using definition of the Sobolev class. This yields the wished upper bound

$$\sup_{f^{X,Y}\in\mathcal{S}(\alpha,Q)}\mathbb{E}\{\int_{\mathcal{R}}(\tilde{f}^{X,Y}(x,y) - f^{X,Y}(x,y))^2\} \leq B_* n^{-1}J^2 + QJ^{-2\alpha} \leq (B_* + Q)n^{-\alpha/(\alpha+1)}. \tag{3}$$

Now let us explain how $\tilde{\theta}_{ji}$, satisfying (2), can be constructed. Given Assumptions 1 and 3, we can write for the likelihood (mixed joint density) of the observed quintet,

$$f^{AV,AW,\Delta,\Gamma,A}(x,y,1,1,1) = kf^{X,Y}(x,y)S^{C,D}(x,y). \tag{4}$$

This formula allows us to write down the Fourier coefficient θ_{ji} as the expectation,

$$\begin{aligned}\theta_{ji} &= \int_{\mathcal{R}} f^{X,Y}(x,y)\varphi_{ji}(x,y)dydx = \int_{\mathcal{R}} \frac{f^{AV,AW,\Delta,\Gamma,A}(x,y,1,1,1)}{kS^{C,D}(x,y)}\varphi_{ji}(x,y)dydx \\ &= \mathbb{E}\left\{\frac{A\Delta\Gamma\varphi_{ji}(V,W)I((V,W)\in\mathcal{R})}{kS^{C,D}(V,W)}\right\}. \end{aligned}\tag{5}$$

Formula (5) suggests the following sample-mean estimator,

$$\hat{\theta}_{ji} = \frac{1}{n}\sum_{l=1}^{n}\frac{A_l\Delta_l\Gamma_l\varphi_{ji}(V_l,W_l)I((V_l,W_l)\in\mathcal{R})}{kS^{C,D}(V_l,W_l)}. \tag{6}$$

To use (6) we need to estimate k and $S^{C,D}(x,y)$. The availability likelihood k is estimated by the sample mean $\hat{k} = \frac{1}{n}\sum_{i=1}^{n} A_i$. Estimation of $S^{C,D}(x,y)$ is more involved and is based on the following formula,

$$\begin{aligned}\hat{S}^{C,D}(x,y) = \frac{1}{n} + \exp\Bigg(&-\sum_{l=1}^{n}\frac{A_l(1-\Delta_l)I(A_lV_l\le x)}{\sum_{i=1}^{n}I(A_iV_i\ge A_lV_l)} \\ &-\sum_{l=1}^{n}\frac{A_l(1-\Gamma_l)I(A_lW_l\le y)I(A_lV_l>x)}{1+\sum_{i=1}^{n}I(A_iV_i\ge x)I(A_iW_i\ge A_lW_l)}\Bigg).\end{aligned}\tag{7}$$

Now we can plug these estimates in (6) and get the data-driven Fourier coefficient estimate

$$\tilde{\theta}_{ji} = \sum_{l=1}^{n}\frac{A_l\Delta_l\Gamma_l\varphi_{ji}(A_lV_l,A_lW_l)I((A_lV_l,A_lW_l)\in\mathcal{R})}{\hat{S}^{C,D}(A_lV_l,A_lW_l)\sum_{r=1}^{n}A_r}. \tag{8}$$

It is possible to show that this coefficient estimate satisfies (2). This yields the following theorem.

Theorem 1. *Suppose that Assumptions 1–3 hold and parameter α is known. Then the bivariate-density estimate (1) with Fourier estimates (8) is rate optimal, namely it satisfies (3).*

In practice the parameter α is either specified (typically $\alpha = 1$ or $\alpha = 2$ are chosen), or adaptive estimator is used. Following [7], the next section proposes an adaptive estimator for small samples.

3 Estimator for Small Samples

The estimator depends on four non-negative parameters c_{TH}, c_{JM}, c_{J0}, and c_{J1}.
Step 1. Set $J_n = \lfloor c_{J0}+c_{J1}\ln(n)\rfloor, c_{J0} > 0, c_{J1} > 0$ and $J_M = J_n * c_{JM}$. Calculate the sample mean Fourier coefficient estimates $\tilde{\theta}_{ji}$ for $j, i = 0, 1, \ldots, J_M$ using (8). Then calculate the empirical variances of $\tilde{\theta}_{ji}$ and denote them as $\tilde{v}_{jin}$.
Step 2. Calculate the empirical cutoffs $\tilde{J}_1$ and $\tilde{J}2$ that minimize the empirical risk

$$(\tilde{J}_1, \tilde{J}_2) := \text{argmin}_{0\le J_1, J_2 \le J_n} \sum_{j=0}^{J_1}\sum_{i=0}^{J_2}(2\tilde{v}_{jin} - \tilde{\theta}^2_{ji}). \tag{9}$$

Step 3. Calculate the series estimate

$$\hat{f}^{X,Y}(x,y) = \sum_{j=0}^{\tilde{J}_1}\sum_{i=0}^{\tilde{J}_2}\tilde{\theta}_{ji} I(\tilde{\theta}^2_{ji} > c_{TH}\tilde{v}_{jin})\varphi_{ji}(x,y). \tag{10}$$

Step 4. If the estimate takes on negative values, use the L_2-projection on the class of nonnegative densities. Similarly, if it is known that the density is monotone, use the L_2-projection on monotonic densities. The projections are defined and discussed in [3].

4 Simulations and Real Data Analysis

4.1 Numerical Study

Let us present a numerical study that sheds some light on the performance of the proposed estimator. To study the effect of smoothness of the underlying function, we generate the underlying sample (X, Y) from smoother normal and rougher strata distributions, defined in [3]. Censoring random variables (C, D) are independently generated from Uniform(0,2) and Exponential(rate=1/1.5). For underlying normal distribution, $C, D \sim U(0, 2)$ yield 45% censoring and $C, D \sim Exponential(rate = 1/1.5)$ yield 50% censoring. For underlying strata distribution, $C, D \sim U(0, 2)$ yield 40% censoring and $C, D \sim Exponential(rate = 1/1.5)$ yield 45% censoring. For each combination of degree of missingness and censoring, 100 simulations were ran and the integrated squared error was averaged. In Tables 1 and 2, AISEM, AISENM and AISEU denote average integrated squared error for estimators based on missing censored data, censored data with no missing and underlying uncensored data with no missing, respectively.

For both normal and strata underlying densities, censoring affects the performance of the estimator more than missing. Indeed, the ratio of AISE for missing censored data and underlying data is much larger than that for missing censored data and censored data without missing. For smoother normal distribution, censoring affects the quality of estimation more than it does for rougher strata density. For normal density on Table 1, the effects of missing in comparison to only censoring without missing is more pronounced when missingness exceeds

Table 1. Simulation results for density estimation, $X, Y \sim$ normal

	n = 100		n = 200		n = 400	
$C, D \sim$ independent $U(0,2)$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$
$k = 0.95$	1.68	1.01	1.83	1.05	2.27	1.04
$k = 0.9$	1.67	1.00	1.93	1.11	2.45	1.12
$k = 0.85$	1.76	1.06	2.08	1.20	2.68	1.23
$k = 0.8$	1.80	1.08	2.19	1.26	3.04	1.39
$k = 0.75$	1.82	1.09	2.24	1.29	3.27	1.50
$C, D \sim$ independent exponential ($rate = 1/1.5$)						
$k = 0.95$	1.67	1.09	2.08	1.02	2.63	1.03
$k = 0.9$	1.70	1.12	2.17	1.07	2.70	1.05
$k = 0.85$	1.81	1.19	2.28	1.12	2.96	1.16
$k = 0.8$	1.88	1.24	2.43	1.19	3.19	1.25
$k = 0.75$	1.94	1.27	2.72	1.34	3.32	1.30

Table 2. Simulation results for density estimation, $X, Y \sim$ Strata

	n = 100		n = 200		n = 400	
$C, D \sim$ independent $U(0,2)$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$	$\frac{AISEM}{AISEU}$	$\frac{AISEM}{AISENM}$
$k = 0.95$	1.44	1.01	1.45	1.02	1.21	1.03
$k = 0.9$	1.52	1.07	1.50	1.06	1.26	1.06
$k = 0.85$	1.57	1.10	1.56	1.10	1.28	1.09
$k = 0.8$	1.58	1.11	1.59	1.12	1.37	1.16
$k = 0.75$	1.66	1.17	1.69	1.19	1.41	1.19
$C, D \sim$ independent Exponential($rate = 1/1.5$)						
$k = 0.95$	1.58	1.04	1.52	1.05	1.29	1.02
$k = 0.9$	1.59	1.06	1.61	1.11	1.35	1.06
$k = 0.85$	1.64	1.09	1.62	1.11	1.37	1.08
$k = 0.8$	1.67	1.11	1.73	1.19	1.43	1.13
$k = 0.75$	1.75	1.16	1.76	1.21	1.46	1.15

10%; otherwise, the performance of two estimators with censored data are comparable. Estimation for rougher strata density is less affected by missing than smoother normal density. In Table 2, the ratios of AISEM/AISENM are quite close to 1, especially when $k \leq 0.85$. Even for higher degrees of missingness, the performance of two estimators with censored data are comparable. Thus, we conclude that effect of missing is less pronounced for rougher densities.

4.2 Rotterdam Breast Cancer Data

We now analyze a breast cancer dataset which is conveniently located in survival R package and contains time to relapse or last follow-up, time to death or last follow-up and the indicator of censoring for each of these times for 2982 patients. The rate of censoring is 49% for time to relapse and 57% for time to death. The time is in days, and we will construct the bivariate density of the two time-to-event variables. There is no missingness in the data, so we will simulate MCAR missingness and compare the resulting density estimator to that of the full data.

Figure 1 shows the breast cancer data and the 3-dimensional bivariate density estimate of the two time-to-event variables. The interval of estimation is $[0, 5500] \times [0, 6300]$ because no uncensored observations occur outside of that rectangle. Next, we simulate MCAR missingness, compute the estimator of bivariate density and compare it to the benchmark estimator based on full data.

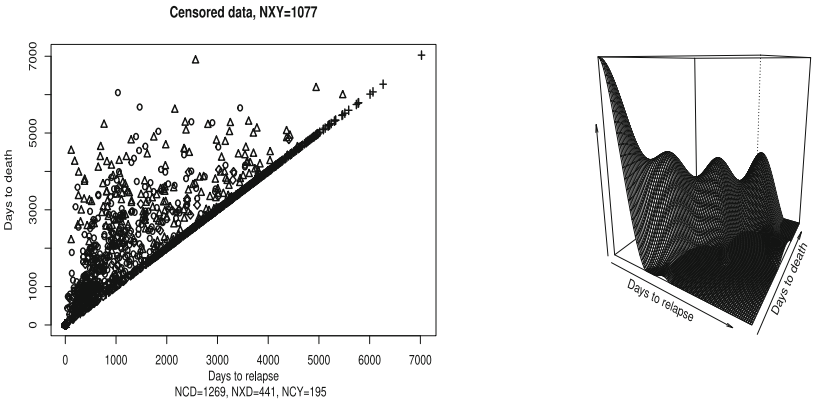


Fig. 1. Bivariate density estimation for all $n = 2982$ observations of breast cancer data. The left diagram shows the censored data. The circles show uncensored observations (X, Y), meaning that $\Delta = 1$ and $\Gamma = 1$ and NXY refers to number of uncensored pairs of observations. The crosses, triangles and rhombuses show the pairs (C, D), (X, D) and (C, Y). The right diagram is the proposed bivariate density estimate based on the full data without missingness.

Figure 2 shows the slices for fixed values of x, which are times to relapse, for a particular simulation of missingness. We observe that the slices are slowly shifting to the left, meaning that longer time to relapse corresponds to longer time to death. We also notice that our proposed estimators for 20% and 25% missingness perform very well for all slices and are very close to the benchmark estimator computed on data without missingness.

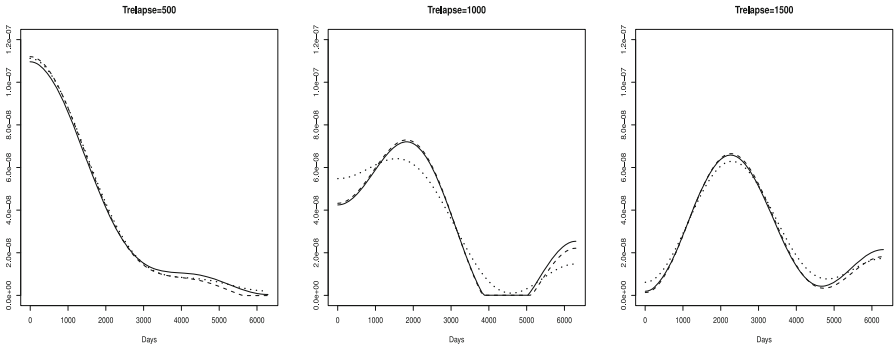


Fig. 2. Slices of bivariate density estimate for breast cancer data where times to relapse are 500, 1000 and 1500 days. The solid line is the benchmark density estimate based on the full data without missingness, the dashed line is estimate when rate of missing is 20%, the dotted line is estimate when rate of missing is 25%.

References

1. Akritas, M.G., Van Keilegom, I.: Estimation of bivariate and marginal distributions with censored data. J. R. Stat. Soc. Ser. B Stat Methodol. **65**, 457–471 (2003)
2. Dabrowska, D.M.: Kaplan-Meier estimate on the plane. Ann. Statist. **16**, 1475–1489 (1988)
3. Efromovich, S.: Nonparametric Curve Estimation: Methods, Theory, and Applications. Springer, New York (1999)
4. Efromovich, S.: On sharp adaptive estimation of multivariate curves. Math. Methods Statist. **9**, 117–139 (2000)
5. Efromovich, S.: Density estimation under random censorship and order restrictions: From asymptotic to small samples. J. Amer. Statist. Assoc. **96**, 667–684 (2001)
6. Efromovich, S.: Adaptive nonparametric density estimation with missing observations. J. Stat. Plan. Inference **143**, 637–650 (2013)
7. Efromovich, S.: Missing and Modified Data in Nonparametric Estimation. CRC Press, Boca Raton, FL (2018)
8. Efromovich, S.: Nonparametric bivariate density estimation for censored lifetimes. Ann. Statist. **50**, 2767–2792 (2022)
9. Hoffmann, M., Lepski, O.: Random rates in anisotropic regression. Ann. Statist. **30**, 325–396 (2002)
10. Hougaard, P.: Analysis of Multivariate Survival Data. Springer, New York (2000)
11. Prentice, R.L., Zhao, S.: Nonparametric estimation of the multivariate survivor function: the multivariate Kaplan-Meier estimator. Lifetime Data Anal. **24**, 3–27 (2018)
12. Rubin, D.B.: Inference and missing data. Biometrika **63**, 581–90 (1976)
13. Tsybakov, A.: Introduction to Nonparametric Estimation. Springer, New York (2009)
14. van Buuren, S.: Flexible Imputation of Missing Data. Chapman and Hall/CRC (2018)

Causal Markov Categories and Possibility Theory

Tobias Fritz[1(✉)] and Pedro Terán[2]

[1] Universität Innsbruck, Innsbruck, Austria
tobias.fritz@uibk.ac.at
[2] Universidad de Oviedo, Oviedo, Spain

Abstract. Markov categories are a recent approach to probability theory that is category-theoretical rather than measure-theoretical. It takes (an abstract version of) Markov kernels, rather than measurable mappings, as primitive. Our aim is to bring researchers in possibility theory and other uncertainty formalism into contact with Markov categories by explaining how the categorical formalism applies in this context as well. We note that, under any continuous triangular norm, possibilistic transition matrices between finite sets form a causal Markov category, just like probabilistic transition matrices do.

Keywords: Markov category · Possibility measure · Possibilistic transition matrix · Triangular norm

1 Introduction

Category theory [13,14] abstracts the commonalities underlying different mathematical structures and aspires to provide a unified account of them with broadly applicable definitions and theorems. Similar to how group theory offers an abstract language to speak about the symmetries of different objects, category theory offers an abstract language to speak about different structures like topological spaces, groups or deductive systems. For example, a homomorphism linking two groups, a continuous function linking two topological spaces, and a proof linking two formulas are all abstracted into the notion of a *morphism* linking two *objects* in a category.

To get some idea of what a category-theoretical approach to probability might look like, consider the example of *zero–one laws*. These are statements showing that, under suitable assumptions involving independence or conditioning, the probability of a certain event is 0 or 1. But a category-theoretic approach need not be concerned with calculating numerical probability values at all. Instead, it is perfectly enough to have a categorical definition of when a morphism is 'deterministic' and to formulate zero–one laws as sufficient criteria for a given morphism to be deterministic. In particular, [5, Theorems 5.3 and 5.4] states and proves categorical versions of the zero–one laws of Kolmogorov and Hewitt–Savage in this spirit.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 142–148, 2024.
https://doi.org/10.1007/978-3-031-65993-5_17

A brief overview of categorical probability is available in [3, Sect. 1]. Markov categories were introduced by Golubtsov [8], rediscovered by Cho and Jacobs [2], and given their current name in [3]. As already noted, the morphisms in a Markov category are abstract versions of Markov kernels. A substantial number of results from classical probability theory and theoretical statistics have since been generalized to the categorical setting. To name a few, this includes theorems on sufficient statistics [3], the de Finetti criterion [6] and the d-separation criterion for Bayesian networks [7].

The fact that numerical probabilities and their concrete properties are not used in this approach leads to the question to which extent this level of abstraction can also cover numerical valuations of uncertainty which do not obey the probability axioms, e.g., possibility theory, imprecise probabilities, and so on. In possibility theory, the role of the sum is played by the maximum, and some possibilistic versions of the kind of theorems mentioned above, like the law of large numbers, do exist (e.g., [15]). A hypothetical categorical proof of the law of large numbers might unify both versions.

In this paper, a much more modest objective is set. We first all want to advertise Markov categories as a general framework for theories of uncertainty. Second, we will show that one of the categorical properties abstracted from probability—the so-called *causality axiom*—also holds in the possibilistic setting.

Providing self-contained definitions of a Markov category and a causal Markov category for readers without prior knowledge of category theory has not been possible for space reasons, and moreover those definitions are not used in the proofs; hence we chose to avoid them and introduced intuitive explanations instead. The proofs themselves are fully rigorous.

2 Preliminaries

The minimum and maximum will be denoted by $\wedge$ and $\vee$, respectively. A *triangular norm* $\top$, often called a *t-norm* for short, is a binary operation on $[0,1]$ which is associative, commutative, has 1 as its neutral element, and is increasing as a bivariate function. There follows $x \top 0 = 0$ for all $x \in [0,1]$. A t-norm is called *continuous* if it is continuous as a bivariate function.

A *possibility measure* is a set function Π on the power set $\mathcal{P}(\Omega)$ of an underlying set Ω such that $\Pi(\emptyset) = 0, \Pi(\Omega) = 1$ and $\Pi(\bigcup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i)$ for any set family $(A_i)_{i \in I}$. Domains with extra structure can be considered, but since this paper deals with finite sets, they will not be relevant here. In the finite case, the third condition is equivalent to $\Pi(A \cup B) = \Pi(A) \vee \Pi(B)$. Thus Π is uniquely identified by the *possibility distribution* $\pi : \omega \in \Omega \mapsto \pi(\omega) = \Pi(\{\omega\})$.

A *category* consists of a collection of *objects* and a collection of *morphisms* which go 'from' a source object 'to' a target object, hence the notation $f : X \to Y$. Morphisms, though, can be things other than functions. A morphism $f : X \to Y$ can be composed with a morphism $g : Y \to Z$, resulting in a specified morphism $g \circ f : X \to Z$. Composition is associative. Each object X has a morphism id_X such that, for any morphism $f : X \to Y$, both $f \circ \mathrm{id}_X$ and $\mathrm{id}_Y \circ f$ equal f.

The formal definition of a *Markov category*, which space reasons prevent us from unpacking here, is discussed in [3, Sect. 2]. For our purposes, the reader can think of an object as a set and a morphism as a (generalization of a) Markov kernel, i.e., it assigns to each input value in X a (generalization of a) probability distribution over the output values in Y. The categories in the next section are very faithful to that idea, transposing it to the realm of possibility instead of probability.

At the intuitive level, things one can do in a Markov category include: creating products $X \otimes Y$ and $f \otimes g$, swapping $X \otimes Y$ into $Y \otimes X$, deleting the information in X, and creating an additional copy in $X \otimes X$ of said information. See [3, Example 2.5] for a description of how this works in the Markov category $\mathsf{FinStoch}$, having finite sets as objects and Markov kernels between them, which can be represented by probabilistic transition matrices, as morphisms.

3 The Markov Category $\mathsf{Kl}(D_{([0,1],\vee,\top)})$

Let $(R, +, \cdot)$ be a semiring. In [4, Example 1.4], the Markov category $\mathsf{Kl}(D_R)$ is considered. Its objects are the finite sets and its morphisms are R-valued transition matrices between finite sets. This generalizes ordinary transition matrices $[p(y|x)]_{x,y}$, which are composed as

$$(q \circ p)(z|x) = \sum_y q(z|y) \cdot p(y|x),$$

by just replacing the sum and product with the semiring operations. Recall that, in a transition matrix, $p(y|x)$ describes the probability of a system evolving to state $y \in Y$ when it starts out in state $x \in X$. The notation Kl reflects the fact that this Markov category is essentially[1] the *Kleisli category* of the monad of R-valued distributions on the category of sets.

Observe that $[0, 1]$, equipped with the maximum $\vee$ as the sum and a triangular norm $\top$ as the product, becomes a commutative semiring. All properties are checked easily from those of t-norms; in particular, distributivity

$$(x \vee y) \top z = (x \top z) \vee (y \top z)$$

is equivalent to the property that a t-norm is increasing as a bivariate function (for more on semirings in possibility theory, see [16] in this volume). Therefore we already know:

Proposition 1. *Let $\top$ be a triangular norm. Then the category $\mathsf{Kl}(D_{([0,1],\vee,\top)})$ having finite sets as objects and possibilistic transition matrices $\pi : Y \times X \to$*

[1] The only difference is that the Kleisli category also contains infinite sets as objects, and a general morphism is defined as a transition matrix with finitely supported columns. We limit ourselves to the finite case for simplicity, while noting that all of our statements apply just as well to the whole Kleisli category.

$[0,1]$ *as morphisms, with* $\max_y \pi(y|x) = 1$ *for all* $x \in X$*, and composition of* $\pi : Y \times X \to [0,1]$ *and* $\mu : Z \times Y \to [0,1]$ *given by*

$$(\mu \circ \pi)(z|x) = \max_{y \in Y} \mu(z|y) \top \pi(y|x),$$

is a Markov category.

Notice that each $\pi(\cdot|x)$ and $\mu(\cdot|y)$, as well as $(\mu \circ \pi)(\cdot|x)$, is a possibility distribution. The composition law in Proposition 1 (analogous to the Chapman–Kolmogorov equation) can be derived, from a prior notion of conditioning, as a theorem on possibilistic Markov processes, see [9, Cor. 15] where $\top$ is any weakly invertible t-norm which distributes over arbitrary suprema. See also [10, p. 245], where $\top$ is the product; and [1], where $\top$ is the minimum.

4 The Causality Axiom

The *causality axiom* is an information flow axiom proposed in [3, Def. 11.31]. Intuitively, it means that if the outcome of a future choice between applying one morphism or another does not correlate with the present, then it cannot correlate with the past either. Thus future information still undetectable in the present cannot 'spill into the past'. For a formal definition and its discussion, see [3, Sect. 11] and [4, Sect. 2.4].

Our main result (Theorem 3) is based on the following property of continuous t-norms which is of independent interest.

Theorem 1. *Let* $\top$ *be a continuous triangular norm. Then, whenever* $s, t, w \in [0,1]$*,*

$$s \top w = t \top w \quad \Rightarrow \quad s \top v = t \top v \ \ \forall v \in [0, w]. \tag{1}$$

Proof. By symmetry, we may assume $s \leq t$. Moreover, the result is trivial if $w = 0$, so we may assume $w > 0$ if needed.

Step 1: $\top$ *is the minimum.* Assume $s \wedge w = t \wedge w$. If $s, t \leq w$ then $s = t$, and if $s, t > w$ then also $s, t > v$. In both cases, the conclusion follows. Lastly, if $s \leq w < t$ then $s = w$. Since $v \leq w$, we have $v \leq s < t$ whence $s \wedge v = t \wedge v$.

Step 2: $\top$ *is the product.* Clearly, $sw = tw$ with $w > 0$ implies $s = t$ and thus $sv = tv$.

Step 3: $\top$ *is the Łukasiewicz t-norm.* That t-norm is given by $x \top y = \max\{x + y - 1, 0\}$. There are two cases. If $s \top w = t \top w = 0$, by the monotonicity of $\top$ also $s \top v = t \top v = 0$. And if $s \top w = t \top w > 0$ then $s + w - 1 = t + w - 1$ whence $s = t$ and the conclusion follows.

Step 4: $\top$ *is an arbitrary continuous t-norm.* The structure of all continuous t-norms is known (see, e.g., [11,12]). Namely, there exists a family $\{[a_i, b_i]\}_{i \in I}$ of subintervals (with non-empty non-overlapping interiors) covering $[0,1]$ and t-norms $\{\top_i\}_{i \in I}$ such that:

(a) Each $\top_i$ is the minimum, the product, or the Łukasiewicz t-norm,

(b) $x \top y = \phi_i(x) \top_i \phi_i(y)$ for some increasing bijective $\phi_i : [a_i, b_i] \to [0,1]$, whenever $x, y \in [a_i, b_i]$ for some $i \in I$,
(c) $x \top y = x \wedge y$ whenever $x, y \notin (a_i, b_i)$ for any $i \in I$.

Fix $w > 0$. For some $i, j \in I$ we have $s \in [a_i, b_i]$ and $t \in [a_j, b_j]$.

Case 4.1: For all $v \in [0, w]$ for which $v \notin (a_i, b_i \wedge w)$ and $v \notin (a_j, b_j \wedge w)$, by Step 1,

$$s \top v = s \wedge v = t \wedge v = t \top v.$$

Case 4.2: Assume now $v \in (a_j, b_j \wedge w)$ and $i \neq j$. Thus $s \leq a_j \leq t$. By Case 4.1, we have $s \top (b_j \wedge w) = t \top (b_j \wedge w)$. The monotonicity of $\top$ implies

$$s \top (b_j \wedge w) = a_j \top (b_j \wedge w) = t \top (b_j \wedge w)$$

whence

$$\phi_j(a_j) \top_j \phi_j(b_j \wedge w) = \phi_j(t) \top_j \phi_j(b_j \wedge w).$$

Thus

$$s \top v = s \wedge v = a_j \wedge v = a_j \top v = \phi_j(a_j) \top_j \phi_j(v) = \phi_j(t) \top_j \phi_j(v) = t \top v$$

where the second identity uses Step 1 and the second-to-last one uses Steps 1–3.

Case 4.3: Assume now $v \in (a_i, b_i \wedge w)$ and $i \neq j$. Thus $t \geq b_i \geq s$. By Cases 4.1 and 4.2, we have $t \top (b_i \wedge w) = s \top (b_i \wedge w)$. The monotonicity of $\top$ implies

$$t \top (b_i \wedge w) = b_i \top (b_i \wedge w) = s \top (b_i \wedge w)$$

whence

$$\phi_i(b_i) \top_i \phi_i(b_i \wedge w) = \phi_i(s) \top_i \phi_i(b_i \wedge w).$$

Thus

$$t \top v = t \wedge v = b_i \wedge v = b_i \top v = \phi_i(b_i) \top_i \phi_i(v) = \phi_i(s) \top_i \phi_i(v) = s \top v$$

where the second identity uses Step 1 and the second-to-last one uses Steps 1–3.

Case 4.4: The remaining case that $i = j$ and $v \in (a_i, b_i \wedge w)$ is similar but $\top_i$ is used instead of $\wedge$.

The relevance of that property is that it characterizes the causality of the Markov category $\mathsf{Kl}(D_{([0,1],\vee,\top)})$.

Theorem 2. *Let $\top$ be a triangular norm. Then $\mathsf{Kl}(D_{([0,1],\vee,\top)})$ is causal if and only if (1) holds.*

The main ingredient in the proof below is the following result [4, Proposition 2.20].

Proposition 2. *Let $(R, +, \cdot)$ be a commutative semiring. Then the Markov category $\mathsf{Kl}(D_R)$ is causal if and only if*

$$s \cdot v = t \cdot v, \quad s \cdot w = t \cdot w$$

whenever $s, t, v, w \in R$ and the following conditions are met:

(i) There exist $x, y, z \in R$ *such that* $s + x = 1_R$, $t + y = 1_R$, $v + w + z = 1_R$;
(ii) $s \cdot (v + w) = t \cdot (v + w)$.

Proof. (of Theorem 2) Since $([0,1], \vee, \top)$ is a commutative semiring, Proposition 2 applies. Moreover, $1_R = 1$. In $[0,1]$, condition (i) always holds by just taking $x = y = z = 1$. Thus we need to check that

$$s \top (v \vee w) = t \top (v \vee w) \quad \Rightarrow \quad s \top v = t \top v,\ s \top w = t \top w \tag{2}$$

is equivalent to (1).

Since $[0,1]$ is totally ordered, $v \vee w$ is either v or w. That means one conclusion in (2) holds trivially and only the other need be checked. By symmetry, assume $v \leq w$ whence (2) simplifies as

$$v \leq w,\ s \top w = t \top w \quad \Rightarrow \quad s \top v = t \top v,$$

which is equivalent to (1).

Therefore, Theorems 1 and 2 establish that finite sets and possibilistic transition matrices between them form a causal Markov category, provided the triangular norm in Proposition 1 is continuous.

Theorem 3. *Let* $\top$ *be a continuous triangular norm. Then the Markov category* $\mathsf{Kl}(D_{([0,1],\vee,\top)})$ *is causal.*

Acknowledgements. Research in this paper was supported by the Austrian Science Fund (FWF P-35992) as well as Spain (PID2019-104486GB-I00) and the Principality of Asturias (SV-PA-21-AYUD/2021/50897).

References

1. Avranchenkov, K., Sanchez, E.: Fuzzy Markov chains and decision-making. Fuzzy Optim. Dec. Mak. **1**, 143–159 (2002)
2. Cho, K., Jacobs, B.: Disintegration and Bayesian inversion via string diagrams. Math. Struct. Comput. Sci. **29**, 938–971 (2019)
3. Fritz, T.: A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. Adv. Math. **370**(107239), 1–105 (2020)
4. Fritz, T., Gonda, T., Houghton-Larsen, N.G., Lorenzin, A., Perrone, P., Stein, D.: Dilations and information flow axioms in categorical probability. Math. Struct. Comput. Sci. **33**, 913–0957 (2023)
5. Fritz, T., Rischel, E.F.: Infinite products and zero-one laws in categorical probability. Compositionality **2**(3), 1–20 (2020)
6. Fritz, T., Gonda, T., Perrone, P.: De Finetti's Theorem in Categorical Probability. J. Stoch. Anal. **2**(4), 6 (2021)
7. Fritz, T., Klingler, A.: The d-separation criterion in categorical probability. J. Mach. Learn. Res. **24**(46), 1–49 (2023)
8. Golubtsov, P.V.: Axiomatic description of categories of information converters. Probl. Inform. Trans. **35**, 259–274 (1999)

9. Janssen, H., de Cooman, G., Kerre, E.E.: First results for a mathematical theory of possibilistic Markov processes. In: 6th International Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp 1425–1432 (1996)
10. Janssen, H.J., de Cooman, G., Kerre, E.E.: Coherence of Dempster's conditioning rule in discrete possibilistic Markov models. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **8**, 241–252 (2000)
11. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer, Dordrecht (2000)
12. Klement, E.P., Mesiar, R., Pap, E.: Archimedean components of triangular norms. J. Aust. Math. Soc. **78**, 239–255 (2005)
13. MacLane, S.: Categories for the Working Mathematician. Springer, New York (1971)
14. Riehl, E.: Category Theory in Context. Dover, Mineola (2016)
15. Terán, P.: Law of large numbers for the possibilistic mean value. Fuzzy Sets Syst. **245**, 116–124 (2014)
16. Terán, P.: Possibilistic variables and Akian semirings. In this volume

Sensitivity Analysis on the Choice of the Metric on Cronbach's α Coefficient for Interval-Valued Data in Questionnaires

José García-García[1(✉)], María Ángeles Gil[1,2], and María Asunción Lubiano[1]

[1] Department of Statistics and Operational Research and Mathematics Didactics, University of Oviedo, 33007 Oviedo, Spain
{garciagarjose,magil,lubiano}@uniovi.es
[2] Spanish Royal Academy of Sciences, 28004 Madrid, Spain

Abstract. Interval-valued scales have become an efficient alternative to conventional single-point scales for capturing richer information in questionnaires measuring imprecise human traits. Thus, new statistical techniques are being developed to analyze this type of data. In this respect, when items in a construct allow respondents to make use of interval-valued scales, the internal consistency reliability can be quantified by means of an extension of the well-known Cronbach α coefficient based on the family of θ-metrics between intervals. Since theoretical results are shown not to be informative enough, the impact of metric choice is discussed through a Monte Carlo study concerning different scenarios of a simulation procedure mimicking human rating.

Keywords: Interval-valued scale · Interval-valued data · θ-distance · Reliability · Internal consistency · Cronbach's α coefficient

1 Introduction

Nowadays, the measurement of attitudinal and behavioral human traits plays a crucial role in decision making in a wide range of fields. For this reason, researchers from different disciplines have started to develop suitable tools to obtain accurate data in connection with people's opinions, interests, emotions, and many health-related features. This kind of latent variables, usually called *constructs*, are typically measured through *questionnaires* involving several *items*.

The choice of an appropriate response type is fundamental. Traditionally, respondents have been allowed to choose the single number, label or symbol that best fits their assessment (agreement, satisfaction, frequency, etc.) within

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 149–157, 2024.
https://doi.org/10.1007/978-3-031-65993-5_18

Likert-type [11] or *visual analogue* [10] *scales*. Despite their well-known advantages, these psychometric instruments do not simultaneously capture the individual differences and the genuine imprecision of human ratings. To overcome this limitation, the so-called *interval-valued scales* [15,17] allow respondents to select a range of values in accordance to their valuation. In recent times, these scales have received an increasing interest due to the development of new statistical techniques for interval-valued data analysis (see, for instance, [14] and references therein) and users general acceptance shown in several studies [5,6].

Once a questionnaire has been conducted, constructs scores are often obtained by adding the collected responses to the involved items. To ensure the reliability of scores derived in this way, the *internal consistency* of the items in each construct, that is, the extent to which respondents have been coherent when rating each underlying dimension, must be checked. *Cronbach's α coefficient* [3] is probably the most popular reliability index for single-point data and its definition can be extended to the interval-valued setting by considering a certain family of metrics as it has been done before for general fuzzy-valued data [13].

The aim of this work is to study the impact of changing the metric on such an internal consistency measure for interval-valued data. Section 2 recalls the necessary tools to define in Sect. 3 the extended Cronbach α coefficient for this type of data. Finally, the influence of the choice of the metric is analyzed in Sect. 4 and some concluding remarks are contained in Sect. 5.

2 Notation and Preliminaries

Let $\mathcal{K}_c(\mathbb{R})$ denote the class of nonempty compact intervals of the space of real numbers $\mathbb{R}$. Any interval $A \in \mathcal{K}_c(\mathbb{R})$ can be written either as $A = [\inf A, \sup A]$, where $\inf A$ and $\sup A$ denote its *infimum* and *supremum*, respectively, or $A = [\operatorname{mid} A - \operatorname{spr} A, \operatorname{mid} A + \operatorname{spr} A]$ or, in short, $A = [\operatorname{mid} A \pm \operatorname{spr} A]$, where $\operatorname{mid} A$ is its *mid-point* or *center* and $\operatorname{spr} A$ is its *spread* or *radius*. This space endowed with the usual interval arithmetic defined by the Minkowski sum, settled as

$$A + B = [\inf A + \inf B, \sup A + \sup B] = [(\operatorname{mid} A + \operatorname{mid} B) \pm (\operatorname{spr} A + \operatorname{spr} B)]$$

where $A, B \in \mathcal{K}_c(\mathbb{R})$, and a product by a scalar operation, given by

$$\lambda \cdot A = \begin{cases} [\lambda \cdot \inf A, \lambda \cdot \sup A] \text{ if } \lambda \geq 0 \\ [\lambda \cdot \sup A, \lambda \cdot \inf A] \text{ if } \lambda < 0 \end{cases} = [\lambda \cdot \operatorname{mid} A \mp |\lambda| \cdot \operatorname{spr} A]$$

for $A \in \mathcal{K}_c(\mathbb{R})$ and $\lambda \in \mathbb{R}$, is not linear but semilinear with conical structure due to the lack of inverse element with respect to Minkowski sum. Hence, there does not exist an always well-defined concept of difference for interval-valued data.

Sometimes, this drawback can be avoided by considering certain families of metrics between intervals. Hereinafter, the *θ-distance* suggested by Gil et al. [9], which is defined for any $A, B \in \mathcal{K}_c(\mathbb{R})$ and $\theta \in (0, \infty)$ as the real number

$$d_\theta(A, B) = \sqrt{(\operatorname{mid} A - \operatorname{mid} B)^2 + \theta \cdot (\operatorname{spr} A - \operatorname{spr} B)^2},$$

is to be considered. Although θ can take any positive value, it is often imposed that $\theta \leq 1$ in order to weight the deviation in location (mid-points) no less than the deviation in imprecision (spreads), as well as to make d_θ coincide with Bertoluzza et al. metric [2]. The most common choices of θ are $\theta = 1/3$, which assigns the same relevance to distances between all the points in the considered intervals, and $\theta = 1$, which is equivalent to weighting only and uniformly the distances between their endpoints and, consequently, makes d_θ coincide with Vitale's ρ_2 metric [16].

Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $\mathcal{X}\colon \Omega \to \mathcal{K}_c(\mathbb{R})$ is said to be a *random interval* if, and only if, both mappings $\inf \mathcal{X}\colon \Omega \to \mathbb{R}$ and $\sup \mathcal{X}\colon \Omega \to \mathbb{R}$, where $(\inf \mathcal{X})(\omega) = \inf \mathcal{X}(\omega)$ and $(\sup \mathcal{X})(\omega) = \sup \mathcal{X}(\omega)$ for all $\omega \in \Omega$, or, alternatively, $\operatorname{mid} \mathcal{X}\colon \Omega \to \mathbb{R}$ and $\operatorname{spr} \mathcal{X}\colon \Omega \to [0, \infty)$, which are defined in an analogous way, are real-valued random variables with respect to $(\Omega, \mathcal{A}, P)$. Moreover, $\mathcal{X}$ is a random interval if it is measurable with respect to $\mathcal{A}$ and the Borel σ-field generated by the topology induced by d_θ on $\mathcal{K}_c(\mathbb{R})$. Therefore, some stochastic concepts such as the (induced) distribution of a random interval or the independence between random intervals can be immediately established.

The statistical behavior of random intervals can be resumed by means of certain summary measures. In particular, given a random interval, $\mathcal{X}$, its central tendency is usually reported through the *Aumann-type mean* value [1], which is defined (whenever the involved expectations exist) as

$$\mathrm{E}(\mathcal{X}) = [E(\inf \mathcal{X}), E(\sup \mathcal{X})] = [E(\operatorname{mid} \mathcal{X}) \pm E(\operatorname{spr} \mathcal{X})],$$

and the *Fréchet-type variance* [7], given by the real number

$$\operatorname{Var}_\theta(\mathcal{X}) = E\left(d_\theta^2(\mathcal{X}, E(\mathcal{X}))\right) = \operatorname{Var}(\operatorname{mid} \mathcal{X}) + \theta \cdot \operatorname{Var}(\operatorname{spr} \mathcal{X}),$$

is used to quantify its deviation. Furthermore, given a simple random sample of size n from $\mathcal{X}$, $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, that is, n independent random intervals that are identically distributed as $\mathcal{X}$, the sample counterpart of the Aumann mean is

$$\overline{\mathcal{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i = \left[\overline{\inf \mathcal{X}}, \overline{\sup \mathcal{X}}\right] = \left[\overline{\operatorname{mid} \mathcal{X}} \pm \overline{\operatorname{spr} \mathcal{X}}\right],$$

and the sample estimate of the Fréchet-type variance is defined as

$${}_\theta S_{\mathcal{X}}^2 = \frac{1}{n} \sum_{i=1}^{n} d_\theta^2\left(\mathcal{X}_i, \overline{\mathcal{X}}\right) = S_{\operatorname{mid} \mathcal{X}}^2 + \theta \cdot S_{\operatorname{spr} \mathcal{X}}^2.$$

3 The Cronbach α Coefficient for Interval-Valued Data

Consider a construct with $k > 1$ items in a questionnaire being conducted on a sample of n respondents. Assume that responses to items are rated through interval-valued scales, so the responses to each item can be modeled as a random interval, $\mathcal{X}_j$ $(j = 1, \ldots, k)$, and, consequently, construct's total score is given by the random interval $\mathcal{X}_{\text{total}} = \mathcal{X}_1 + \cdots + \mathcal{X}_k$. Besides, the available sample

information can be modeled then by means of simple random samples from these random intervals.

The population *Cronbach α coefficient* for interval-valued data and its sample counterpart are given, respectively, whenever the involved Fréchet-type variances exist and the construct's total score is nondegenerate, by the numerical values

$$\alpha_\theta = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \mathrm{Var}_\theta(\mathcal{X}_j)}{\mathrm{Var}_\theta(\mathcal{X}_{\text{total}})}\right) \qquad \text{and} \qquad \widehat{\alpha}_\theta = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} {}_\theta S^2_{\mathcal{X}_j}}{{}_\theta S^2_{\mathcal{X}_{\text{total}}}}\right).$$

This coefficient is always lower than or equal to 1 and the closer to 1, the greater the internal consistency. Actually, it is often considered that values over 0.9 are excellent for individual decision making and values over 0.8 and 0.7 are good and acceptable, respectively, for research purposes in most fields of application.

4 Sensitivity Analysis on the Choice of the Metric

The expression of Cronbach's α coefficient depends on the choice of θ through the distance involved in Fréchet-type variances. In order to analyze the effect of the choice of θ on this coefficient, a theoretical analysis of Cronbach's α coefficient derivatives as functions of θ can be done, yielding the following result.

Proposition 1. *Cronbach's α coefficient for interval-valued data is a monotonic –either increasing or decreasing– function of θ.*

Since this result is shown to be inconclusive, in the sense that the extent of the increase or decrease of Cronbach's α coefficient as function of θ is not quantified in general, the impact of varying the choice of the metric is discussed by considering a Monte Carlo study concerning some representative scenarios of a simulation approach mimicking human rating.

4.1 Simulation Procedure

Interval-valued data for this empirical analysis have been generated by adapting the simulation strategy suggested by de la Rosa de Sáa et al. [4] and taking into account a real-life questionnaire where interval-valued scales were used [12] to fit the involved constraints.

More concretely, once the number of respondents, n, and the number of items, k, are fixed, the answers to the items in a construct of a questionnaire involving interval-valued scales whose reference interval is $[l, u]$ are simulated as follows.

1. An auxiliary sample with n intervals defined in $[0, 1]$, $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$, is generated by considering a mixture of three different procedures.
   - 5% of the auxiliary data is generated by fitting each interval's endpoints to an ordered sample from a beta $\beta\,(p, q)$ distribution.

- 35% of that sample is created in such a way that the mid-points of the intervals, X, follow a $\beta(p, q)$ distribution while their spreads are generated from a uniform $\mathcal{U}(0, \min\{1/3, X, 1 - X\})$ distribution.
- 65% of the interval-valued data of the auxiliary sample is obtained with their mid-points coming from a $\beta(p, q)$ distribution, and their spreads coming from a gamma $\gamma(2, 20)$ distribution.

2. To replicate the foreseeable high correlation between each respondent's answers, for the j-th item ($j = 1, \ldots, k$) two random values, a_j and b_j, are generated from a $\mathcal{U}(0, 1)$ distribution and a normal $\mathcal{N}(.10, .05)$ distribution, respectively. Then, for each $i = 1, \ldots, n$ and each $j = 1, \ldots, k$ an interval based on the auxiliary sample is obtained as $\mathsf{x}_{ij} = a_j \cdot \mathsf{x}_i + b_j + \varepsilon_{ij}$, where ε_{ij} is drawn from a $\mathcal{N}(.10, .05)$ distribution, truncating (if necessary) its extremes to be contained within the interval $[0, 1]$.
3. The intervals are moved to the reference interval of the interval-valued scale considered. Therefore, the answer of the i-th respondent ($i = 1, \ldots, n$) to the j-th item ($j = 1, \ldots, k$) is computed as $\widetilde{\mathsf{x}}_{ij} = (u - l) \cdot \mathsf{x}_{ij} + l$.

This simulation procedure is implemented in R environment through the `simulIVS()` function of the IntervalQuestionStat package [8], which provides several tools to statistically analyze interval-valued responses in questionnaires.

4.2 Results

To ease the graphical display and the associated interpretation, instead of discussing the influence of the choice of the metric as a function of θ, the alternative weight $\tau = \theta/(\theta + 1)$ is considered. Note that, since $\theta \in (0, \infty)$, then $\tau \in (0, 1)$ and values of $\tau \leq .5$ are equivalent to values of $\theta \leq 1$.

In particular, 20 values of τ equally spaced between .05 and .95 are fixed and Cronbach's α coefficient is approximated at each of these points by Monte Carlo method with 100,000 replicates. Thus, for each τ, Cronbach's α coefficient is computed as the arithmetic mean of the estimates of this reliability measure obtained for 100,000 simulated databases containing interval-valued responses.

Moreover, the influence of the choice of θ on Cronbach's α coefficient is analyzed by considering different $\beta(p, q)$ distributions in the simulation strategy described in Sect. 4.1 in order to handle different types of responses (uniform or majority of one or two-tailed extreme ratings) and changing the number of respondents to the questionnaire and the number of items in the construct.

Varying Beta Distribution. Figure 1 shows the simulation results obtained for a fixed number of items and respondents ($k = 10$ and $n = 100$) and assuming different $\beta(p, q)$ distributions. The observed values of Cronbach's α coefficient are higher for symmetric beta distributions although they always report excellent levels of internal consistency. As a function of τ, this coefficient remains relatively stable, especially for values of $\tau \leq .5$ (which corresponds to $\theta \in (0, 1]$). A more noticeable increasing trend could be highlighted when non-symmetric beta distributions are considered but the largest changes observed, apart from

occurring within the interval (.5,1), do not change the interpretation associated with the obtained Cronbach α coefficients.

Fig 1

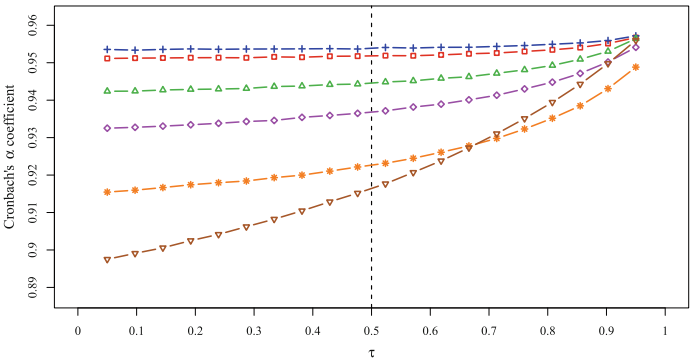


Fig. 1. Cronbach's α coefficient of a construct with 10 interval-valued scale-based items in a questionnaire with 100 respondents, and assuming $\beta(1,1)$ (□), $\beta(.75,.75)$ (+), $\beta(2,2)$ (△), $\beta(4,2)$ (◇), $\beta(6,1)$ (✱), and $\beta(6,10)$ (▽) distributions for simulations

Varying Number of Respondents. The simulated values of Cronbach's α coefficient obtained for a fixed beta distribution ($p = q = 1$), a fixed number of items ($k = 10$), and different numbers of respondents to the questionnaire are represented in Fig. 2. The outputs show that the effect of the number of respondents on the internal consistency of the items in the construct is not so relevant since there is hardly any difference between the observed values of the extended Cronbach α coefficient according to the number of respondents. In this case, excellent values of internal consistency are always attained. Although it may appear from the graph that there is an increasing trend and that Cronbach's α coefficient is therefore influenced by the choice of τ, especially for values of $\tau > .5$, vertical axis values immediately reveals that this changes are almost irrelevant and it remains stable over the interval (0,1).

Fig 2

Varying Number of Items. In Fig. 3, simulation results obtained for a fixed beta distribution ($p = q = 1$), a fixed number of respondents ($n = 100$), and different numbers of items in the construct are represented. The outputs reveal that the greater the number of items, the greater the value of Cronbach's α coefficient, and, thus, the higher the internal consistency. In this scenario, constructs with 5 or more items show excellent reliability. However, Cronbach's α coefficients of constructs with only 2 or 3 items are lower, yielding just acceptable and good consistency values, respectively. As a function of τ, this coefficient remains

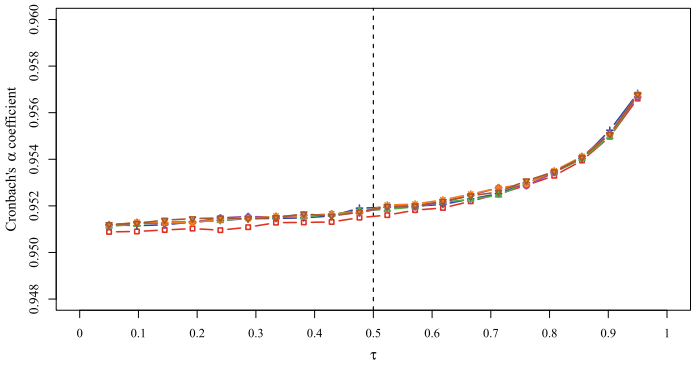


Fig. 2. Cronbach's α coefficient of a construct with 10 interval-valued scale-based items in a questionnaire with 50 (□), 100 (+), 150 (△), 200 (◇), 250 (✱), and 300 (▽) respondents, and assuming a $\beta(1, 1)$ distribution for simulations

relatively stable across (0, 1), although for the smaller number of items, a slight increasing trend becomes noticeable from $\tau = .8$ (which is equivalent to $\theta = 4$) onwards but, in any case, the choice of τ seems to be almost irrelevant.

Fig 3

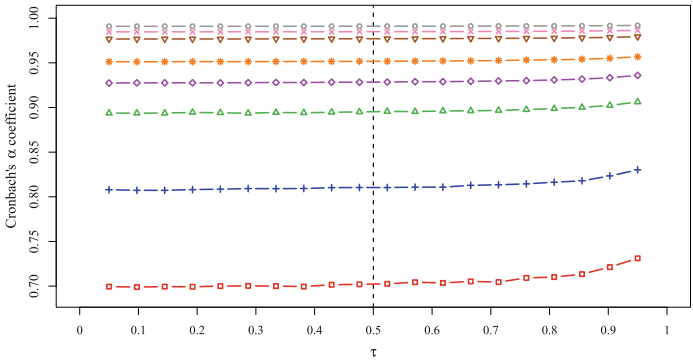


Fig. 3. Cronbach's α coefficient of a construct with 2 (□), 3 (+), 5 (△), 7 (◇), 10 (✱), 20 (▽), 30 (×), and 50 (○) interval-valued rating scale-based items in a questionnaire with 100 respondents, and assuming a $\beta(1, 1)$ distribution for simulations

5 Concluding Remarks

Cronbach's α coefficient definition can be extended in order to measure the internal consistency reliability of interval-valued items in a construct for a given performance of a questionnaire rating imprecise human traits. Such an extension

involves the family of θ-metrics between intervals, so the value of Cronbach's α coefficient for interval-valued data depends on the choice of θ.

This work shows that metric choice, that is, the choice of θ, scarcely affects the value of Cronbach's α coefficient for interval-valued data, especially for values of $\theta \leq 1$, which are the most often used in practice, independently of response type, number of respondents to the questionnaire, and the number of items in a construct. Therefore, the choice of the metric is almost irrelevant to this purpose.

Acknowledgements. This research has been partially supported by Spanish Ministry of Science, Innovation and Universities Grant PID2019-104486GB-I00, the Principality of Asturias Grant AYUD/2021/50897 as well as the Severo Ochoa—Principality of Asturias Grant BP22-173 (García-García). Their financial support is gratefully acknowledged. The authors would like also to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of this work.

References

1. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. **12**(1), 1–12 (1965). https://doi.org/10.1016/0022-247X(65)90049-1
2. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathw. Soft Comput. **2**, 71–84 (1995)
3. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika **16**(3), 297–334 (1951). https://doi.org/10.1007/BF02310555
4. de la Rosa de Sáa, S., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A.: Fuzzy rating scale-based questionnaires and their statistical analysis. IEEE Trans. Fuzzy Syst. **23**, 111–126 (2015). https://doi.org/10.1109/TFUZZ.2014.2307895
5. Ellerby, Z., Wagner, C.: Do people prefer to give interval-valued or point estimates and why? In: 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). pp. 1–6 (2021). https://doi.org/10.1109/FUZZ45933.2021.9494507
6. Ellerby, Z., Wagner, C., Broomell, S.B.: Capturing richer information: on establishing the validity of an interval-valued survey response mode. Behav. Res. Methods **54**, 1240–1262 (2022). https://doi.org/10.3758/s13428-021-01635-0
7. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. Ann. Inst. Henri Poincaré **10**(4), 215–310 (1948)
8. García-García, J.: IntervalQuestionStat: Tools to Deal with Interval-Valued Responses in Questionnaires (2022), https://CRAN.R-project.org/package=IntervalQuestionStat, R package version 0.2.0
9. Gil, M.Á., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. Metrika **56**, 97–111 (2002). https://doi.org/10.1007/s001840100160
10. Hayes, M.H.S., Patterson, D.G.: Experimental development of the graphic rating method. Psychol. Bull. **18**, 98–99 (1921)
11. Likert, R.: A technique for the measurement of attitudes. Arch. Psychol. **22**, 140–155 (1932)
12. Lubiano, M.A., García-García, J., García-Izquierdo, A.L., Castaño, A.M.: The extended version of Cohen's d index for interval-valued data. In: García-Escudero, L.A. et al. (eds.) Building Bridges between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing, vol. 1433, pp. 263–270. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-15509-3_35

13. Lubiano, M.A., García-Izquierdo, A.L., Gil, M.Á.: Fuzzy rating scales: does internal consistency of a measurement scale benefit from coping with imprecision and individual differences in psychological rating? Inf. Sci. **550**, 91–108 (2021). https://doi.org/10.1016/j.ins.2020.10.042
14. Sinova, B., Van Aelst, S.: A spatial-type interval-valued median for random intervals. Statistics **52**(3), 479–502 (2018). https://doi.org/10.1080/02331888.2018.1435658
15. Themistocleous, C., Pagiaslis, A., Smith, A., Wagner, C.: A comparison of scale attributes between interval-valued and semantic differential scales. Int. J. Mark. Res. **61**(4), 394–407 (2019). https://doi.org/10.1177/1470785319831227
16. Vitale, R.A.: L^p metrics for compact, convex sets. J. Approx. Theory **45**(3), 280–287 (1985). https://doi.org/10.1016/0021-9045(85)90051-6
17. Wagner, C., Miller, S., Garibaldi, J.M., Anderson, D.T., Havens, T.C.: From interval-valued data to general type-2 fuzzy sets. IEEE Trans. Fuzzy Syst. **23**(2), 248–269 (2015). https://doi.org/10.1109/TFUZZ.2014.2310734

On a Linear Regression Model for Star-Shaped Set-Valued Random Elements

José Graña-Colubi[1], Gil González-Rodríguez[2], and Ana Belén Ramos-Guajardo[2(✉)]

[1] Department of Statistics, University of Oviedo, OR and MD, Spain
[2] INDUROT/Department of Statistics, University of Oviedo, OR and MD, Spain
{gil,ramosana}@uniovi.es

Abstract. The aim of this work is to propose a linear regression model between two random elements that take values in a particular star-shaped sets space. Random star-shaped sets will be introduced, as well as the corresponding framework to deal with them. This framework relies on a center-radial characterization, leveraging the well-established properties of Hilbert spaces. The key features of the proposed linear regression model will be analyzed.

Keywords: Star-shaped set · Set-valued data · Linear regression · Random element

1 Introduction

During last decades, numerous statistical methodologies for handling random variables taking imprecise values as outputs have been emerged (see [5–8,11,15] among others). From a theoretical standpoint, random variables that associate the results of a random experiment with elements in general spaces are considered. Notably, this includes instances such as random intervals, which are random elements taking values in the space of real compact intervals (see, for instance, [2–4,17,18]).

Random (fuzzy) sets appear as extensions of random intervals to a higher dimensional spaces dealing with elements belonging to the space of compact and convex (fuzzy) subsets of $\mathbb{R}^p$. Extensively explored in existing literature (consult [1,9,10,13,14,16] among others), these random elements have garnered significant attention and analysis.

In addressing interval data, there is a broad consensus in the scientific community to use Minkowski's arithmetic. The extension of Minkowski arithmetic to the manipulation of compact and convex subsets within the d-dimensional Euclidean space, along with its fuzzy counterpart, coincides with the arithmetic induced by the conventional one in functional spaces, particular when the support function is used to characterize these elements [12,20]. Nevertheless, it is

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 158–165, 2024.
https://doi.org/10.1007/978-3-031-65993-5_19

well known that employing the support function poses challenges in terms of intuitiveness, particularly when attempting to establish relationships between the support function of aggregated sets and the sets themselves, especially in dimensions greater than 1. This difficulty arises primarily due to the convexification property inherent in Minkowski arithmetic, as discussed extensively in [12]. Consequently, an alternative generalization of Minkowski's arithmetic is proposed in [12] which leads to the use of the so-called star-shaped sets [19] and their characterization. This approach involves defining a center point within the object, serving as a location indicator, alongside a radial function capturing directional inaccuracies from this center.

This contribution presents a regression model that establishes a linear relationship between two star-shaped random sets. This model extends the regression framework discussed in [2] for interval-valued data. Besides, the well-definition of the model is checked and some key properties are analyzed.

The manuscript is structured as follows. In Sect. 2, some preliminaries on star-shaped sets in $\mathbb{R}^d$ ($p \geq 1$) are provided. Section 3 is devoted to the introduction of the random star-shaped sets and their associated main moments. The regression model proposed for random star-shaped sets, along with its principal properties, is examined in Sect. 4. Finally, Sect. 5 synthesizes the conclusions drawn from the study and outlines potential open problems for future research.

2 Preliminaries on Star-Shaped Sets

Star-shaped sets can be seen as a higher-dimensional extension of intervals. Let $d \in \mathbb{N}$, $A \subset \mathbb{R}^d$ is called a star-shaped set if, and only if, there exists $x \in A$ so that

$$\lambda x + (1-\lambda)a \subset A \ \forall \lambda \in [0,1] \text{ and } \forall a \in A \ . \tag{1}$$

The methodology outlined in [12] will be followed, wherein a systematic approach for managing these sets is introduced and their principal properties are studied. Thus, given $A \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$, it is defined

$$A + x = \{a + x \mid a \in A\} \text{ and } A - x = \{a - x \mid a \in A\} \ .$$

If A^* is star-shaped, the set of $x \in A^*$ that verify (1) is named *kernel* of A^*, and it is denoted by $ker(A^*)$. A^* can be expressed through a tuple (a, A), for a given $a \in ker(A)$ and $A = A^* - a$, particularly $0 \in ker(A)$ (see [12] for more details about this characterization).

For all $A \subset \mathbb{R}^d$ so that $0 \in A$, its associated polar function $\rho_A : \mathbb{S}^{d-1} \to \overline{\mathbb{R}}$ is defined as

$$\rho_A(u) = \sup\{\lambda \in [0, \infty] \mid \lambda u \in A\} \ \forall u \in \mathbb{S}^{d-1}$$

where $\mathbb{S}^{d-1}$ represents the unit sphere in $\mathbb{R}^d$.

In the same way, given a function $f : \mathbb{S}^{d-1} \to \mathbb{R}$ its associated star-shaped set is defined in [12] as

$$K_f = \{\lambda u \mid \lambda \in [0, \infty),\ \lambda \leq |f(u)|,\ u \in \mathbb{S}^{d-1}\} \ .$$

For the construction of the aforementioned framework, star-shaped sets must also verify certain properties (see [12]). Tuples (a, A) will henceforth be employed, where $a \in \mathbb{R}^d$ and $A \subset \mathbb{R}^d$ denotes star-shaped set such that $0 \in ker(A)$ and $\rho_A \in \mathcal{L}_2(\mathbb{S}^{d-1}, \lambda_{\mathbb{S}^{d-1}})$. The space composed of equivalence classes of these sets is denoted by $\mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$, where the equivalence classes in $\mathbb{X}_0^2(\mathbb{R}^d)$ are induced by those in $\mathcal{L}_2(\mathbb{S}^{d-1}, \lambda_{\mathbb{S}^{d-1}})$. This notation offers the advantage of distinct interpretation for each component: the first one represents the location of the element whereas the second is related to the shape and the size which, in certain contexts, can be interpreted as the imprecision of the element.

2.1 Properties

Firstly, an arithmetic on $\mathbb{X}^2(\mathbb{R}^d) = \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$ is introduced in [12]. Namely, given (a, A), $(b, B) \in \mathbb{X}^2(\mathbb{R}^d)$, it is established that

$$\begin{aligned}(a, A) + (b, B) &= (a + b, K_{\rho_A + \rho_B}) \\ \lambda(a, A) &= (\lambda a, K_{|\lambda| \rho_A (sign(\lambda)\cdot)}) \ .\end{aligned}$$

The existence of a symmetric element with respect to the addition is not guaranteed. Nonetheless, in specific scenarios, an extension of *Hukuhara's difference* can be considered. That is, Hukuhara's difference of (a, A) and (b, B), denoted by $(a, A) \ -_H \ (b, B)$, is defined as the element $(c, C) \in \mathbb{X}^2(\mathbb{R}^d)$ such that $(a, A) = (b, B) + (c, C)$, if it exists.

Furthermore, a metric for the space $\mathbb{X}^2(\mathbb{R}^d)$ is defined in [12] as follows,

$$d_\tau((a, A), (b, B)) = \left((1 - \tau)\|a - b\|^2 + \tau \|\rho_A - \rho_B\|_2^2\right)^{1/2} \ ,$$

where $\|\cdot\|$ represents the usual norm in R^d and $\|\cdot\|_2$ the usual norm in $\mathcal{L}_2(\mathbb{S}^{d-1}, \lambda_{\mathbb{S}^{d-1}})$. The parameter $\tau \in (0, 1)$ enables the weighting of the relative significance between the distance of locations and the distance of shapes.

Consider now the vector space $(\mathcal{H}_2, +, \cdot)$, where $\mathcal{H}_2 = \mathbb{R}^d \times \mathcal{L}_2(\mathbb{S}^{d-1}, \lambda_{\mathbb{S}^{d-1}})$. By taking $\langle \cdot, \cdot \rangle_\tau$ as

$$\langle (x_1, f_1), (x_2, f_2) \rangle_\tau = (1 - \tau)\langle x_1, x_2 \rangle + \tau \int f_1 f_2 d\lambda_{\mathbb{S}^{d-1}}$$

the space $(\mathcal{H}_2, +, \cdot, \langle \cdot, \cdot \rangle_\tau)$ becomes a separable Hilbert space. The mapping $\Gamma : \mathbb{X}^2(\mathbb{R}^d) \to \mathcal{H}_2$ given by $\Gamma(a, A) = (a, \rho_A)$ $\forall (a, A) \in \mathbb{X}^2(\mathbb{R}^d)$ is an isometric embedding that preserves the arithmetic formerly introduced ([12]).

3 Random Star-Shaped Sets

In this section, the concept of *random $\mathcal{L}_2$ star-shaped set* and some associated properties are recalled from [12]. These random elements can be used to model random experiments for which the characteristic of interest is expressed through star-shaped sets.

Definition 1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A mapping $X : \Omega \to \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$ is called* random $\mathcal{L}_2$ *star-shaped set if it is a Borel measurable mapping with respect to $\mathcal{A}$ and the Borel $\sigma-$field generated by the topology induced by the metric d_τ in $\mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$.*

Given $X : \Omega \to \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$, whenever it is needed, the marginal mappings are denoted as $m_X : \Omega \to \mathbb{R}^d$ and $S_X : \Omega \to \mathbb{X}_0^2(\mathbb{R}^d)$, referring to the random location counterpart and the random star-shaped part respectively. For short $\rho_X : \Omega \to \mathcal{L}_2(\mathbb{S}^{d-1})$ will denote the associated polar mapping. Then $\Gamma(X) = \Gamma(m_X, S_X) = (m_X, \rho_X)$. Particularly, X is a random $\mathcal{L}_2$ star-shaped set if, and only if, m_X, S_X and ρ_X are random elements. Consequently, being a random $\mathcal{L}_2$ star-shaped set does not depend on the value of $\tau \in (0, 1)$ chosen.

Definition 2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random $\mathcal{L}_2$ star-shaped set so that $E(\|m_X\|) < \infty$ and $E(\|\rho_X\|_2) < \infty$. The* expected value of X *is defined as the element $E(X) \in \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$ such that $\Gamma(E(X)) = E(\Gamma(X))$, where the expectations in $\mathcal{H}_2$ are considered in terms of Bochner integrals.*

Note that, in particular, $E(X) = (E(m_X), E(S_X)) = (E(m_X), K_{E(\rho_X)})$.

Definition 3. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random $\mathcal{L}_2$ star-shaped set so that $E(\|m_X\|^2) < \infty$ and $E(\|\rho_X\|_2^2) < \infty$ (equivalently, $E(d_\tau^2(X, 0)) < \infty$), the* variance of X *is defined as*

$$\sigma_X^2 = Var(X) = E\left(d_\tau^2(X, E(X))\right) = E\left(\|\Gamma(X) - E(\Gamma(X))\|_\tau^2\right) = Var(\Gamma(X)) \ .$$

Definition 4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and X, Y two random $\mathcal{L}_2$ star-shaped sets so that $E(d_\tau^2(X, 0)) < \infty$ and $E(d_\tau^2(Y, 0)) < \infty$, the* covariance of X and Y *is defined as*

$$\sigma_{X,Y} = Cov(X, Y) = E\left(\langle \Gamma(X) - E(\Gamma(X)), \Gamma(Y) - E(\Gamma(Y))\rangle_\tau\right) \ .$$

4 Linear Regression Model

Let $(\Omega, \mathcal{A}, P)$ be a probability space and X, Y be two random $\mathcal{L}_2$ star-shaped sets such that $0 < \sigma_{m_X}^2, \sigma_{S_X}^2, \sigma_{m_Y}^2, \sigma_{S_Y}^2 < \infty$, being the variance of m_X and m_Y considered coordinate-wise. The aim is to model the relation between X and Y by means of a concrete error additive regression model of the form

$$\begin{aligned} Y &= f(X) + \varepsilon \ , \\ E(\varepsilon|X) &= (w, W) \in \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d) \end{aligned} \tag{2}$$

being $f : \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d) \to \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d)$ and where the conditional expectation can be introduced as in Def. 2 by considering the corresponding one for Banach spaces (see [21]) through the isometric embedding Γ.

In this particular proposal, for the sake of simplicity, it will be assumed that the components representing location and imprecision are treated independently.

That is, S_Y will be related to S_X and each 1-dimensional component of m_Y will be related to its respective of m_X. Formally,

$$f(X) = (\mathbf{A}m_X, \beta S_X) \tag{3}$$

where

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_d \end{pmatrix} ; \ \alpha_i \in \mathbb{R} \ \forall i = 1, ..., d \tag{4}$$

and

$$0 \leq \beta \in \mathbb{R} \ . \tag{5}$$

In view of some of the properties shown in the previous sections for random $\mathcal{L}_2$ star-shaped sets it is satisfied that for this simple regression model,

$$m_Y = \mathbf{A}m_X + m_\varepsilon \ \ \textit{and} \ \ \rho_Y = \beta\rho_X + \rho_\varepsilon \ .$$

The following result states that, under the imposed restrictions, (2–5) constitute a well-defined model.

Proposition 1. ***(Uniqueness)*** *Let (X, Y) be two random $\mathcal{L}_2$ star-shaped sets satisfying (2) and suppose that, in addition,*

$$Y = (m_Y, S_Y) = (\tilde{\mathbf{A}}m_X, \tilde{\beta}S_X) + \tilde{\varepsilon}$$

with

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{\alpha}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\alpha}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\alpha}_d \end{pmatrix} ; \ \tilde{\alpha}_i \in \mathbb{R} \ \forall i = 1, ..., d \ ,$$

$$0 \leq \tilde{\beta} \in \mathbb{R} \ ,$$

and

$$E(\tilde{\varepsilon}|X) = (\tilde{w}, \tilde{W}) \in \mathbb{R}^d \times \mathbb{X}_0^2(\mathbb{R}^d) \ .$$

It is necessary that $\tilde{\mathbf{A}} = \mathbf{A}$ and $\tilde{\beta} = \beta$.

Proof. Regarding the first component of the model $\tilde{\mathbf{A}}m_X + m_{\tilde{\varepsilon}} = \mathbf{A}m_X + m_\varepsilon$, and consequently $(\tilde{\mathbf{A}} - \mathbf{A})m_X = m_{\tilde{\varepsilon}} - m_\varepsilon$.

By taking the conditional expectation with respect to X it is verified that $(\tilde{\mathbf{A}} - \mathbf{A})m_X = \tilde{w} - w$. Finally, as $\sigma^2_{m_X} > 0$, then the above equality can only hold if $\tilde{\mathbf{A}} - \mathbf{A} = 0$ which proves the first part of the result.

With respect to the second parameter, $\tilde{\beta}S_X + S_{\tilde{\varepsilon}} = S_Y = \beta S_X + S_\varepsilon$. By using the properties of Γ, $S_Y = \beta S_X + S_\varepsilon$ if, an only if,

$$\Gamma(S_Y) = \Gamma(\beta S_X + S_\varepsilon) \Longleftrightarrow \rho_Y = |\beta| \cdot \rho_X(sign(\beta)\cdot) + \rho_\varepsilon \ .$$

As $\beta \geq 0$ (and $\tilde{\beta} \geq 0$), then

$$\tilde{\beta}\rho_X + \rho_{\tilde{\varepsilon}} = \rho_Y = \beta\rho_X + \rho_\varepsilon$$

and, proceeding analogously as in the previous case, it is shown that $\tilde{\beta} = \beta$, which guarantees the uniqueness of the model. □

Note that $\beta \geq 0$ is a necessary condition. In another case, given X such that $\rho_X(u) = \rho_X(-u) \ \ \forall u \in \mathbf{S}^{d-1}$, one can take $\tilde{\beta} = -\beta$ and obtain

$$|\tilde{\beta}| \cdot \rho_X(sign(\tilde{\beta})\cdot) = |\beta| \cdot \rho_X(sign(\beta)\cdot) \ ,$$

which implies that $\tilde{\beta}S_X = \beta S_X$ with $\tilde{\beta} \neq \beta$.

Proposition 2 shows that it is also possible to express the parameters β and α_j, $j \in \{1, \ldots, d\}$, in terms of some population moments of X and Y, in analogy to the classical case.

Proposition 2. *Let (X, Y) be two random $\mathcal{L}_2$ star-shaped sets satisfying (2). Then*

$$\alpha_j = \frac{\sigma_{m_{X_j}, m_{Y_j}}}{\sigma^2_{m_{X_j}}} \ \ \forall j = 1, ..., d \ , \quad \beta = \frac{\sigma_{S_X, S_Y}}{\sigma^2_{S_X}},$$
$$w = E(m_Y) - AE(m_X) \ \text{ and } \ W = E(S_Y) -_H \beta E(S_X) \ .$$

Proof. Given $j \in \{1, ..., d\}$, α_j represents the coefficient of a simple linear regression between m_{Y_j} and m_{X_j}, so its theoretical value is well-known. Furthermore, by applying the properties of Γ,

$$\begin{aligned}\sigma_{S_X, S_Y} &= E\left(\langle \Gamma(S_Y) - \Gamma(E(S_Y)), \Gamma(S_X) - \Gamma(E(S_X))\rangle_\tau\right) \\ &= E\left(\langle \Gamma(\beta S_X + S_\varepsilon) - \Gamma(E(\beta S_X + S_\varepsilon)), \Gamma(S_X) - \Gamma(E(S_X))\rangle_\tau\right) \\ &= \beta\sigma^2_{S_X} + \sigma_{S_\varepsilon, S_X} = \beta\sigma^2_{S_X} \ .\end{aligned}$$

Since $0 < \sigma^2_{S_X} < \infty$,

$$\beta = \frac{\sigma_{S_X, S_Y}}{\sigma^2_{S_X}} \ .$$

Finally,

$$\begin{aligned}E(m_Y) &= E(Am_X + \varepsilon) = AE(m_X) + w, \\ E(S_Y) &= E(\beta S_X + \varepsilon) = \beta E(S_X) + W \ ,\end{aligned}$$

which leads to the result. □

5 Conclusions

A linear regression model between two random star-shaped sets has been presented. The parametrization employed enables the consideration of one linear

relationship between the imprecision of the random sets and a different one between the locations of both of them. This simple model also supposes that each coordinate of the location is solely associated with its corresponding counterpart in the explanatory variable. Furthermore, the theoretical values of the parameters have been derived. This model, although simple, lays the groundwork for future advancements, including potential modifications such as incorporating interactions between mid and spread components or among the various coordinates determining the location of the random elements.

Due to the particularities of the framework, the practical estimation of the parameters from a sample requires to take into account some considerations, such as the existence of Hukuhara's difference. This fact is illustrated by the last proposition, wherein the computation of the intercept for the second part of the model entails the computation of a difference. This means that the use of a least squares approach to solve the estimation problem should take into account this restriction leading to a constrained optimization problem that will be investigated in the near future.

References

1. Aletti, G., Bongiorno, E.: A decomposition theorem for fuzzy set-valued random variables. Fuzzy Set. Syst. **219**, 98–112 (2013)
2. Blanco-Fernández, A., Corral, N., González-Rodríguez, G.: Estimation of a flexible simple linear model for interval data based on set arithmetic. Comput. Stat. Data An. **55**(9), 2568–2578 (2011)
3. Blanco-Fernández, A., González-Rodríguez, G.: Inferential studies for a flexible linear regression model for interval-valued variable. Int. J. Comput. Math. **93**(4), 658–675 (2016)
4. Choirat, C., Seri, R.: Bootstrap confidence sets for the Aumann mean of a random closed set. Comput. Stat. Data An. **71**, 803–817 (2013)
5. Colubi, A., González-Rodríguez, G.: Fuzziness in data analysis: towards accuracy and robustness. Fuzzy Set. Syst. **281**, 260–271 (2015)
6. Couso, I., Dubois, D.: Statistical reasoning with set-valued information: Ontic versus epistemic views. Int. J. Approx. Reason. 55, 1502–1518 (2014)
7. Diamond, P., Kloeden, P.: Characterization of compact subsets of fuzzy sets. Fuzzy Set. Syst. **29**, 341–348 (1989)
8. D'Urso, P., Giordani, P.: A least squares approach to principal component analysis for interval valued data. Chemom. Intell. Lab. Syst. **70**(2), 179–192 (2004)
9. Ferraro, M.B., Coppi, R., González-Rodríguez, G., Colubi, A.: A linear regression model for imprecise response. Int. J. Approx. Reason. **51**(7), 759–770 (2010)
10. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. Fuzzy Set. Syst. **157**, 2608–2613 (2006)
11. Gonzalez-Rodriguez, G., Colubi, A., Gil, M.A.: Fuzzy data treated as functional data: a one-way anova test approach. Comput. Stat. Data An. **56**, 943–955 (2012)
12. González-Rodríguez, G., Ramos-Guajardo, A., Colubi, A., Blanco-Fernández, A.: A new framework for the statistical analysis of set-valued random elements. Int. J. Approx. Reason. **92**, 279–294 (2018)

13. Körner, R.: An asymptotic α-test for the expectation of random fuzzy variables. J. Statist. Plan. Infer. **83**, 331–346 (2000)
14. Molchanov, I.: Theory of Random Sets. Springer, Berlin Heidelberg (2005)
15. Puri, M., Ralescu, D.: Fuzzy random variables. J. Math. Anal. Appl. **114**, 409–422 (1986)
16. Ramos-Guajardo, A.B., Lubiano, M.A.: K-sample tests for equality of variances of random fuzzy sets. Comput. Stat. Data An. **56**(4), 956–966 (2012)
17. Ramos-Guajardo, A.B., Grzegorzewski, P.: Distance-based linear discriminant analysis for interval-valued data. Inf. Sci. **372**, 591–607 (2016)
18. Sinova, B., De La Rosa De Saa, S., Casals, M.R., Gil, M.A., Salas, A.: Interval arithmetic-based simple linear regression between interval data: discussion and sensitivity analysis on the choice of the metric. Inf. Sci. **292**, 347–363 (2016)
19. Toranzos, F.: Radial functions of convex and star-shaped bodies. Am. Math. Mon. **74**, 278–280 (1967)
20. Trutschnig, W., González-Rodríguez, G., Colubi, A., Gil, M.A.: A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Inf. Sci. **179**, 3964–3972 (2009)
21. Vakhania, N., Tarieladze, V., Chobanyan, S.: Probability Distributions on Banach Spaces. Kluwer Academic Publishers Group (1987)

Two-Sample Depth-Based Test for Dispersion in Fuzzy Environment

Przemyslaw Grzegorzewski[1,2(✉)] and Anna Kozak[1]

[1] Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland
{przemyslaw.grzegorzewski,anna.kozak}@pw.edu.pl
[2] Polish Academy of Sciences, Systems Research Institute, Newelska 6, 01-447 Warsaw, Poland

Abstract. The problem of comparing the variability of two populations in the presence of fuzzy data is considered. We show that applying the modified band depth to specifically transformed membership functions, combined with a permutation procedure, leads us to construct a new interesting test for dispersion.

Keywords: functional data analysis · fuzzy data · fuzzy numbers · statistical depth · credibility distribution · test for dispersion

1 Introduction

Disadvantages like problems with subtraction and division of fuzzy numbers, the lack of universally accepted total ranking between fuzzy numbers which precludes rank tests, or a lack of suitable models for the distribution of random fuzzy numbers, disable a direct generalization of classical statistical tests to the fuzzy domain. This is the reason we do not yet have an extensive statistical toolbox for fuzzy data. At the same time, their designers must be inventive in using various ideas to overcome the above-mentioned obstacles.

The literature on tests for comparing variability in fuzzy environment is quite limited. Ramos-Guajardo and Lubiano [8] proposed the bootstrap generalization of the Levene test for random fuzzy sets to examine homoscedasticity of k populations. Grzegorzewski [2] introduced two generalizations of the Sukhatme test for interval-valued data and a two-sample permutation test for fuzzy data [3].

As we know, a convenient tool for multidimensional or functional data analysis is the so-called statistical depth [11]. Recently, three generalizations of simplicial depth to fuzzy sets were proposed by González-de la Fuente et al. [1]. However, since we can identify each fuzzy set with its membership function, a fuzzy sample can be viewed as a set of functions. Therefore, we can try to apply an appropriately chosen depth used in functional data analysis (FDA) to fuzzy data. This very conclusion motivated us for the present contribution. More specifically, we showed that applying the modified band depth to ICr curves (i.e.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 166–174, 2024.
https://doi.org/10.1007/978-3-031-65993-5_20

specifically transformed membership functions), combined with a permutation procedure, led us to construct an interesting test for a two-sample dispersion problem.

The paper is organized as follows: in Sect. 2 we recall basic information on fuzzy number representations. Next, in Sect. 3 we discuss briefly the band depth for functional data and its modified version that seems to suit well a fuzzy environment. In Sect. 4 we introduce the ICr functions which may provide a convenient representation of fuzzy numbers [4] to use some FDA methods. Then, in Sect. 5 we propose a two-sample test for comparing dispersion in the presence of fuzzy data.

2 Fuzzy Data

Let $\mathbb{F}(\mathbb{R})$ denote a space of all fuzzy numbers. As it is known, each fuzzy number has two equivalent representations. Using the **LR-representation**, the membership function μ of $A \in \mathbb{F}(\mathbb{R})$ can be represented in the following form

$$\mu(x) = \begin{cases} L\left(\frac{b-x}{b-a}\right) & \text{if } \ a < x \leqslant b, \\ 1 & \text{if } \ b \leqslant x \leqslant c, \\ R\left(\frac{x-c}{d-c}\right) & \text{if } \ c \leqslant x < d, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $L, R : \mathbb{R} \to [0,1]$ are decreasing functions, called **shape functions**, such that $L(0) = R(0) = 1$, $L(1) = R(1) = 0$, $\forall x > 0$ $L(x), R(x) < 1$ and $\forall x < 1$ $L(x), R(x) > 0$. If $L = R$ and $b-a = d-c$ then a given fuzzy number is symmetric. The most often used fuzzy numbers are **trapezoidal fuzzy numbers** with shape functions $L(x) = R(x) = \max\{0, 1-x\}$ or **triangular fuzzy number**, i.e. trapezoidal fuzzy numbers with $b = c$.

Following the **LU-representation**, a fuzzy number A is completely characterized by a family of its α-cuts $\{A_\alpha\}_{\alpha\in[0,1]}$, where $A_\alpha = \{x \in \mathbb{R} : \mu(x) \geqslant \alpha\}$ if $\alpha \in (0,1]$ and $A_\alpha = cl\{x \in \mathbb{R} : \mu(x) > 0\}$ if $\alpha = 0$ (cl stands for the closure). Thus each α-cut of $A \in \mathbb{F}(\mathbb{R})$ is a nonempty compact interval $A_\alpha = [A_\alpha^L, A_\alpha^U]$, where $A_\alpha^L = \inf A_\alpha$ and $A_\alpha^U = \sup A_\alpha$ denote its lower and upper endpoint.

Since further on we consider fuzzy samples that are realizations of fuzzy random variables, let us also recall the definition of a **random fuzzy number**, also known as a **fuzzy random variable**.

Definition 1. (cf. [7]) *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A mapping $X : \Omega \to \mathbb{F}(\mathbb{R})$ is a random fuzzy number if for all $\alpha \in [0,1]$ the α-level function is a compact random interval.*

3 Statistical Depth for Functional Data

Functional data analysis (FDA) generally refers to situations where the available samples consist of real functions $\mathbb{x} = (x_1, \ldots, x_n)$, where $x_i = x_i(t)$, $i = 1, \ldots, n$,

are defined on a compact interval $\mathcal{I}$. We usually assume that the sample space $\mathcal{X}$ is a real separable Banach space and hence our observations are drawn from an $\mathcal{X}$-valued random element X defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Sometimes a structure of the Hilbert space is assumed for $\mathcal{X}$. Two standard choices for the sample space $\mathcal{X}$ are $\mathcal{C}(\mathcal{I})$, i.e. the Banach space of real continuous functions $x : \mathcal{I} \to \mathbb{R}$ endowed with the supremum norm $||\cdot|| = \sup_t |x(t)|$, or the Hilbert space of square integrable real functions on $\mathcal{I}$ endowed with the usual inner product $\langle x, y\rangle = \int_{\mathcal{I}} x(t)y(t)\,dt$. An important tool used in FDA is a statistical depth that makes some univariate methods based on order statistics possible for functional data. Roughly speaking, a statistical depth allows ordering a sample of functions from the center outward.

López-Pintado and Romo [6] defined the **band depth** by referring to a graph-based approach. Here is, in brief, their idea. Let $x_1(t), \ldots, x_n(t)$ denote a sample of functions in $\mathcal{C}(\mathcal{I})$. A graph of a function x is the subset of the plane $G(x) = \{(t, x(t)) : t \in \mathcal{I}\}$. The band in $\mathbb{R}^2$ delimited by the curves $x_{i_1}, \ldots, x_{i_k}$ is

$$B(x_{i_1}, \ldots, x_{i_k}) = \big\{(t, y(t)) : t \in \mathcal{I},\ \min_{r=1\ldots,k} x_{i_r}(t) \leqslant y(t) \leqslant \max_{r=1\ldots,k} x_{i_r}(t)\big\}.$$

For any function $x \in \{x_1, \ldots, x_n\}$ and any fixed j, where $2 \leqslant j \leqslant n$, the quantity

$$BD_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1\leqslant i_1<i_2<\ldots<i_j\leqslant n} \mathbb{1}\big(G(x) \subseteq B(x_{i_1}, x_{i_2}, \ldots, x_{i_j})\big) \tag{2}$$

is a proportion of bands $B(x_{i_1}, x_{i_2}, \ldots, x_{i_j})$ determined by j different curves $x_{i_1}, x_{i_2}, \ldots, x_{i_j}$ containing the whole graph of x.

Definition 2. *Let J, where $2 \leqslant J \leqslant n$, denote a fixed value. The **band depth** of any curve x in a sample $x_1(t), \ldots, x_n(t)$ is defined by*

$$BD_{n,J}(x) = \sum_{j=2}^{J} BD_n^{(j)}(x). \tag{3}$$

If $X_1(t), \ldots, X_n(t)$ are independent copies of a stochastic process $X(t)$ generating our observations $x_1(t), \ldots, x_n(t)$ then the population version of the band depth (3) for a given curve $x = x(t)$ with respect to the probability measure $\mathbb{P}$ is defined by $BD_J(x, \mathbb{P}) = \sum_{j=2}^{J} BD_n^{(j)}(x, \mathbb{P}) = \sum_{j=2}^{J} \mathbb{P}\big(G(x) \subseteq B(X_1, \ldots, X_j)\big)$, where $B(X_1, \ldots, X_j)$ is a band delimited by j random curves.

If the curves in a sample are very irregular then many of them have the same band depth which results in poor depth-based ranking. Therefore, López-Pintado and Romo [6] proposed the following modified version of (3), which measures the proportion of time that a curve $x(t)$ is lying inside each band.

Definition 3. *Let J, where $2 \leqslant j \leqslant n$, denote a fixed value. The **modified band depth** of any curve x in a sample $x_1(t), \ldots, x_n(t)$ is given by*

$$MBD_{n,J}(x) = \sum_{j=2}^{J} MBD_n^{(j)}(x), \tag{4}$$

where

$$MBD_n^{(j)}(x) = \binom{n}{j}^{-1} \frac{1}{\ell(\mathcal{I})} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant n} \ell\big(\mathcal{I}_B(x; x_{i_1}, x_{i_2}, \ldots, x_{i_j})\big), \tag{5}$$

$\mathcal{I}_B(x; x_{i_1}, \ldots, x_{i_k}) = \big\{t \in \mathcal{I} : \min\limits_{r=1\ldots,k} x_{i_r}(t) \leqslant x(t) \leqslant \max\limits_{r=1\ldots,k} x_{i_r}(t)\big\}$ *and* ℓ *is the Lebesgue measure on* $\mathcal{I}$*.*

One should realize that the computational costs of calculating DB or MBD depend both on the sample size and the number of curves determining a band. Since the order of curves induced by the band depth is very stable in J, its small value is recommended to avoid computational issues.

4 ICr Functions

When considering how to make FDA methods more useful in statistical reasoning based on fuzzy data, we found that another representation of fuzzy numbers than discussed in Sect. 2 would be useful. A representation having properties convenient for FDA methods was suggested in [4] and is based on inverse functions of credibility distributions (cf. [5]).

Suppose $A \in \mathbb{F}(\mathbb{R})$ has a membership function $\mu(x)$ given by (1). Let us extend the sides L and R of A to the real domain as follows

$$L_{ext}(x) = \begin{cases} 0 & \text{if } a < x, \\ L\left(\frac{b-x}{b-a}\right) & \text{if } a \leqslant x < b, \\ 1 & \text{if } b \leqslant x, \end{cases} \qquad R_{ext}(x) = \begin{cases} 1 & \text{if } x \leqslant c, \\ R\left(\frac{x-c}{d-c}\right) & \text{if } c < x \leqslant d, \\ 0 & \text{if } d < x. \end{cases} \tag{6}$$

Obviously, $\mu(x) = L_{ext}(x) - \big[1 - R_{ext}(x)\big]$ for any $x \in \mathbb{R}$.

Definition 4. (Liu [5]) *The* ***credibility distribution*** *of* $A \in \mathbb{F}(\mathbb{R})$ *is a function* $\Upsilon : \mathbb{R} \to [0, 1]$ *defined by*

$$\Upsilon(x) = \frac{1}{2}\Big(L_{ext}(x) + \big[1 - R_{ext}(x)\big]\Big), \quad \forall x \in \mathbb{R}. \tag{7}$$

Looking for a justification for Def. 4 one may notice that Liu [5] defined the credibility distribution as the average of the possibility and necessity functions, i.e. $\Upsilon(x) = \frac{1}{2}\Big(\mathrm{Pos}(x) + \mathrm{Nec}(x)\Big)$, where $\mathrm{Pos}(x) = \sup_{t \leqslant x} \mu(t) = L_{ext}(x)$ and $\mathrm{Nec}(x) = 1 - \sup_{t > x} \mu(t) = 1 - R_{ext}(x)$. It is also worth noting that Stefanini and Guerra [10] considered a more general construction, called the λ-Average Cumulative Function (ACF) $\Psi^{(\lambda)} : \mathbb{R} \to [0, 1]$, defined as follows $\Psi^{(\lambda)}(x) = (1 - \lambda)L_{ext}(x) + \lambda\big[1 - R_{ext}(x)\big]$, where $\lambda \in [0, 1]$. Obviously, $\Upsilon(x) = \Psi^{(\lambda)}(x)$ if and only if $\lambda = 1/2$. Hence, the credibility distribution (7) is a special case of ACF.

Unfortunately, credibility distributions of fuzzy numbers in a sample may have different supports. Meanwhile, the inverse credibility distributions are always defined on the interval $[0, 1]$, which makes them better suited to represent fuzzy numbers in FDA.

Although we could continue the discussion for any fuzzy numbers, in the following, for simplicity, we will limit our considerations to **regular fuzzy numbers**, i.e. fuzzy numbers with continuous membership functions and a single-element core (satisfying $b = c$ in (1)). A family of all regular fuzzy numbers will be denoted by $\mathbb{F}_r(\mathbb{R})$. The following theorem holds.

Theorem 1. *$A \in \mathbb{F}_r(\mathbb{R})$ if and only if its credibility distribution $\Upsilon(x)$ is strictly increasing on $\{x \in \mathbb{R} : 0 < \Upsilon(x) < 1\}$. Moreover, $A \in \mathbb{F}_r(\mathbb{R})$ if and only if it has a unique inverse credibility distribution*

$$\upsilon(\alpha) := \Upsilon^{-1}(\alpha) \tag{8}$$

and $\upsilon(\alpha)$ is continuous and strictly increasing for $\alpha \in [0, 1]$.

Further on the **inverse credibility distribution**, defined by (8), will be abbreviated as the **ICr function**. An example of a regular fuzzy number, its credibility distribution and the corresponding ICr function is given in Fig. 1.

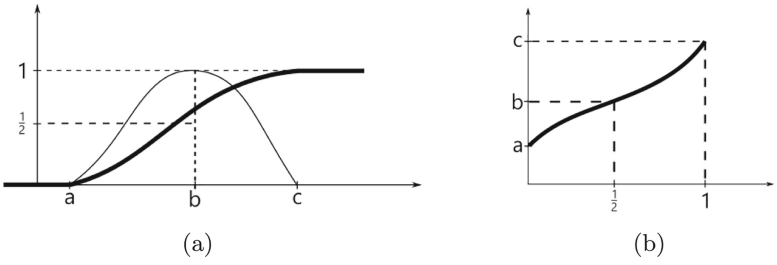


Fig. 1. (a) The membership function of a regular fuzzy number and its credibility distribution $\Upsilon(x)$. (b) The corresponding ICr function $\upsilon(\alpha)$.

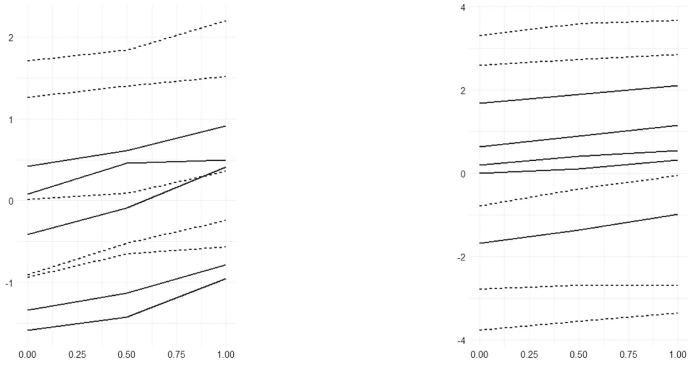
ICr functions provide another representation of fuzzy numbers because there is a one-to-one relationship between membership and ICr functions. They have also some interesting analytical properties, e.g., they are non-decreasing and their common support is the unit interval $[0, 1]$. Moreover, the ICr functions of the regular fuzzy numbers are continuous in $[0, 1]$, and a value of ICr for $1/2$ indicates always the mode of the corresponding membership function.

5 Two-Sample Dispersion Depth-Based Test for Fuzzy Data

Let $\mathbb{X} = (X_1, \ldots, X_n)$ and $\mathbb{Y} = (Y_1, \ldots, Y_m)$ denote two independent random fuzzy samples drawn from populations with unknown distributions. Our goal

is to verify the null hypothesis $H_0 : X \stackrel{d}{=} Y$ that both samples are identically distributed against the alternative hypothesis H_1 that the dispersion of X's and Y's distributions differ significantly.

Let $\mathbb{x} = (x_1, \ldots, x_n)$ and $\mathbb{y} = (y_1, \ldots, y_m)$ denote the realizations of the considered fuzzy samples, where $x_i, y_j \in \mathbb{F}_r(\mathbb{R})$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Before carrying out the test, we must prepare the data appropriately, i.e. transform membership functions of fuzzy observations into their credibility distributions and then to the corresponding ICr functions. More precisely, starting from the following transformations: $\mu_{x_i}(t) \mapsto \Upsilon_{x_i}(t)$ and $\mu_{y_j}(t) \mapsto \Upsilon_{y_j}(t)$, where $t \in \mathbb{R}$, we determine $\upsilon_{x_i}(\alpha) := \Upsilon_{x_i}^{-1}(\alpha)$ and $\upsilon_{y_j}(\alpha) := \Upsilon_{y_j}^{-1}(\alpha)$ for and $i = 1, \ldots, n$, $j = 1, \ldots, m$. As a result we obtain: $\boldsymbol{\upsilon}_x = \big(\upsilon_{x_1}(\alpha), \ldots, \upsilon_{x_n}(\alpha)\big)$ and $\boldsymbol{\upsilon}_y = \big(\upsilon_{y_1}(\alpha), \ldots, \upsilon_{y_m}(\alpha)\big)$, $\alpha \in [0, 1]$. In Fig. 2 we show two experimental situations. In Fig. 2(a) we can see the ICr curves corresponding to samples that do not differ in dispersion, whereas looking at Fig. 2(b) we notice immediately that both samples differ in scale (it seems that the distribution of the sample represented by the solid lines is less dispersed than the second distribution represented by the dashed lines).



(a) (b)

Fig. 2. Two datasets illustrating different arrangements of ICr functions: $\boldsymbol{\upsilon}_x = \big(\upsilon_{x_1}(\alpha), \ldots, \upsilon_{x_n}(\alpha)\big)$ (solid lines) and $\boldsymbol{\upsilon}_y = \big(\upsilon_{y_1}(\alpha), \ldots, \upsilon_{y_m}(\alpha)\big)$ (dashed lines).

Let $\boldsymbol{w} := \boldsymbol{\upsilon}_x \uplus \boldsymbol{\upsilon}_y$, where $\uplus$ stands for vector concatenation pooling two samples into one, i.e. $w_i = \upsilon_{x_i}$ if $1 \leqslant i \leqslant n$ and $w_i = \upsilon_{y_{i-n}}$ if $n + 1 \leqslant i \leqslant N$, where $N = n + m$. Now we can determine MBD for each observation of $\boldsymbol{\upsilon}_x$ with respect to the pooled sample $\boldsymbol{w}$. Keeping in mind that $\mathcal{I} = [0, 1]$ and assuming $J = 2$, for any $\upsilon_{x_i} \in \{\upsilon_{x_1}, \ldots, \upsilon_{x_n}\}$, by (4) and (5), we obtain

$$MBD_N(\upsilon_{x_i}) := MBD_N^{(2)}(\upsilon_{x_i}) = \binom{N}{2}^{-1} \sum_{1 \leqslant i_1 < i_2 \leqslant n} \ell\big(\mathcal{I}_B(\upsilon_{x_i}; w_{i_1}, w_{i_2})\big). \quad (9)$$

Similarly, for each $\upsilon_{y_i} \in \{\upsilon_{y_1}, \ldots, \upsilon_{y_m}\}$ we have

$$MBD_N(\upsilon_{y_i}) := MBD_N^{(2)}(\upsilon_{y_i}) = \binom{N}{2}^{-1} \sum_{1 \leqslant i_1 < i_2 \leqslant n} \ell\big(\mathcal{I}_B(\upsilon_{y_i}; w_{i_1}, w_{i_2}\big). \quad (10)$$

MBD values obtained from (9) and (10) induce a curve order concerning the center of the pooled data set. Hence, MBD values aggregated separately for X and Y samples can provide a basis to conclude whether both samples are approximately equally dispersed or not. Therefore, the test statistic for the proposed two-sample depth-based test for dispersion is defined by

$$T(\boldsymbol{\upsilon}_x, \boldsymbol{\upsilon}_y) = \Big| \sum_{i=1}^{n} MBD_N(\upsilon_{x_i}) - \sum_{i=1}^{m} MBD_N(\upsilon_{y_i}) \Big|. \quad (11)$$

To conclude whether to reject or not the null hypothesis H_0 we determine a p-value, following the approach typical of permutation tests. Let $\boldsymbol{w}^*$ denote a permutation of the initial dataset $\boldsymbol{w}$. Next we take first n elements of $\boldsymbol{w}^*$ and assign them to sample $\boldsymbol{\upsilon}_x^*$, while the remaining m elements create the second sample $\boldsymbol{\upsilon}_y^*$. Roughly speaking, it works like a random assignment of $N = n + m$ elements into two samples of the size n and m, respectively.

Then, using formulas (9)–(10) we calculate the corresponding value t^* of the test statistic (11) for $\boldsymbol{\upsilon}_x^* = \big(\upsilon_{x_1}^*, \ldots, \upsilon_{x_n}^*\big)$ and $\boldsymbol{\upsilon}_y^* = \big(\upsilon_{y_1}, ^* \ldots, \upsilon_{y_m}^*\big)$, i.e. $t^* = T(\boldsymbol{\upsilon}_x^*, \boldsymbol{\upsilon}_y^*)$. Repeating the whole procedure K times we obtain test statistic values t_k^*, for $k = 1, \ldots, K$ permutations, to determine the approximate p-value

$$\text{p-value} = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\big(t_k^* \geqslant t_0\big), \quad (12)$$

where $t_0 = T(\boldsymbol{\upsilon}_x, \boldsymbol{\upsilon}_y)$ is the test statistic value received for the original samples.

6 Simulation Study

To explore the properties of the proposed test we conducted a simulation study. Due to the limited length of the contribution we show only some of its results on situations when both samples $\mathbb{X} = (X_1, \ldots, X_n)$ and $\mathbb{Y} = (Y_1, \ldots, Y_m)$ were triangular random fuzzy numbers. The cores (modes) of $\mathbb{X}$ and $\mathbb{Y}$ observations were generated from the normal distribution $\mathrm{N}(\theta_x, \sigma_x)$ and $\mathrm{N}(\theta_y, \sigma_y)$, respectively, while the distances between the cores and the left and right endpoints of the support were simulated from the uniform distribution.

We examined the stability of the test size (i.e. the supremum of the probability of making a type I error) at 5% significance level. In each test, $K = 1000$ permutations were drawn and the empirical percentages of rejections under H_0 were determined. The results obtained both for equal and nonequal sample sizes showed that the size of our test is stable at the desired level. Then we determined the power of the proposed test. Starting from the situation when both

samples come from the same distribution and then enlarging the difference in dispersion between both samples, we estimated the probability of rejection of the null hypothesis H_0. Figure 3 shows the test power for $n = m = 10$. Experiments performed on samples of different sizes show that the proposed test behaves as expected.

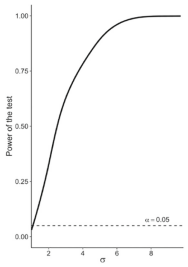


Fig. 3. Power function of the proposed test.

7 Conclusions and Further Research

In this paper, we suggested a new two-sample test for dispersion dedicated to fuzzy data. Since based on the modified band depth it is completely nonparametric which is important in statistical inference based on fuzzy samples. It also utilizes the ICr-based representation of fuzzy numbers which appears promising in further use of FDA methods in a fuzzy environment. Of course, the proposed test requires further in-depth research, including power comparison with other tests (cf. [3,8]) but the preliminary results look promising.

References

1. González-de la Fuente, L., Nieto-Reyes, A., Terán, P.: Simplicial depths for fuzzy random variables. Fuzzy Sets Syst. **471**, 108678 (2023)
2. Grzegorzewski, P.: Two-sample dispersion tests for interval-valued data. In: Medina, J. et al. (Eds.) Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2018), Proceedings, Part III. Communications in Computer and Information Science, vol. 855, pp. 40–51. Springer (2018)
3. Grzegorzewski, P.: Two-sample dispersion problem for fuzzy data. In: Lesot et al. (eds.) Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 82–96. Springer (2020). https://doi.org/10.1007/978-3-030-50153-2_7
4. Grzegorzewski, P., Kozak, A.: Considerations on the use of FDA methods in statistical inference with fuzzy data. In: Proceedings of the IPMU 2024 Conference, submitted (2024)
5. Liu, B.: Uncertainty Theory. Springer (2007)

6. López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Am. Stat. Assoc. **104**, 718–734 (2009)
7. Puri, M.-L., Ralescu, D.-A.: Fuzzy random variables. J. Math. Anal. Appl. **114**, 409–422 (1986)
8. Ramos-Guajardo, A.B., Lubiano, M.A.: K-sample tests for equality of variances of random fuzzy sets. Comput. Stat. Data Anal. **56**, 956–966 (2012)
9. Ramsay, J.-O., Silverman, B.-W.: Functional Data Analysis. Springer (2005)
10. Stefanini, L., Guerra, M.-L.: On possibilistic representations of fuzzy intervals. Inf. Sci. **405**, 33–54 (2017)
11. Zuo, Y., Serfling, R.: General notions of statistical depth function. Ann. Stat. **28**, 461–482 (2000)

Deep Meta Reinforcement Learning for Rapid Adaptation In Linear Markov Decision Processes: Applications to CERN's AWAKE Project

Simon Hirlaender[1(✉)], Sabrina Pochaba[1], Lamminger Lukas[1], Andrea Santamaria Garcia[2], Chenran Xu[2], Jan Kaiser[3], Annika Eichler[3,4], and Verena Kain[5]

[1] IDA Lab Salzburg, Jakob-Haringer-Straße 6, 5020 Salzburg, Austria
simon.hirlaender@plus.ac.at
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
[3] Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[4] Hamburg University of Technology, 21073 Hamburg, Germany
[5] CERN, 1211 Geneva 23, Meyrin, Switzerland

Abstract. Real-world applications of reinforcement learning (RL) face challenges such as the need for numerous interactions and achieving stable training under dynamic conditions. Meta-RL emerges as a solution, particularly in environments where simulations cannot perfectly mimic real-world conditions. This study demonstrates Meta-RL's potential in the CERN's AWAKE project, focusing on the electron line's control. By incorporating Model-Agnostic Meta-Learning (MAML), we showcase how Meta-RL facilitates rapid adaptation to environmental changes with minimal interaction steps. Our findings indicate Meta-RL's efficacy in managing Partially Observable Markov Decision Processes (POMDPs) with evolving hidden parameters, underlining its significance in high-dimensional control challenges prevalent in particle physics experiments and beyond.

Keywords: Meta RL · Accelerator Control · Model-Agnostic Meta-Learning · High-Dimensional Control · Markov Decision Processes · Partial Observability · Machine learning · Artificial intelligence

1 Introduction

Particle accelerators like CERN's Advanced Wakefield Experiment (AWAKE) are crucial for exploring particle physics, presenting unique challenges in beam dynamics control. RL and its advanced form, Meta RL, have emerged as key computational tools for automating control and adapting to dynamic environments in such complex systems. Meta RL, with its "learn to learn" capability,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 175–183, 2024.
https://doi.org/10.1007/978-3-031-65993-5_21

addresses the dimensionality and observability challenges, promising faster adaptation to new tasks and dynamic changes. In this paper, we explore scenarios characterized by a priori partial observability of the system. We employ a simulation to mimic this partial observability through uncertainty, represented by a uniform distribution of unknown parameters. The goal is for the algorithm to adapt to a specific realization within this distribution swiftly. While our scenario is indicative of challenges frequently encountered in accelerator physics control, the methodology is not exclusive to this field. As presumed in our study, problems that start with linearized dynamics can benefit from this approach. This research marks the first application of Meta RL in the domain of accelerator physics control, showcasing its potential to address complex, partially observable systems.

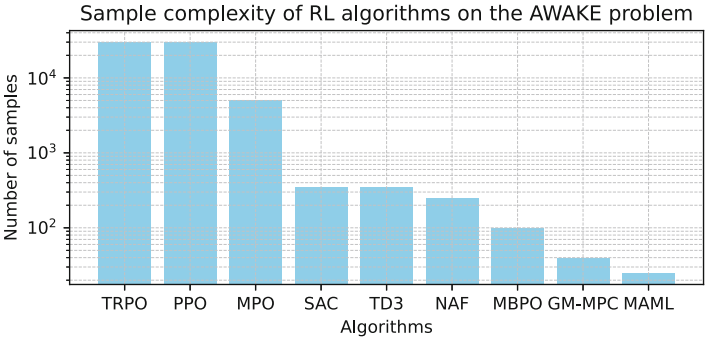


Fig. 1. Comparison of sample complexity across various reinforcement learning algorithms on the AWAKE problem.

2 The AWAKE Environment as a Benchmark

The AWAKE environment is widely recognized as a valuable benchmark for testing RL algorithms, given its relevance to real-world challenges. This has led to its extensive examination within the field [10,12,14,17]. As depicted in Fig. 1, the diversity in sample complexity among various RL algorithms underscores the variability in learning efficiency from environmental interactions.

In our analysis, we assess a spectrum of state-of-the-art RL algorithms applied to the AWAKE problem: Trust Region Policy Optimization (TRPO) [18] and Proximal Policy Optimization (PPO) [20] showcase the stability and ease of use characteristic of on-policy methods, yielding robust performance. Maximum a Posteriori Policy Optimization (MPO), an off-policy gradient-based method [1], and Soft Actor-Critic (SAC), employing a soft Q-learning-based actor-critic framework [7], prioritize exploration via entropy maximization. Twin Delayed Deep Deterministic Policy Gradient (TD3) [4] and Normalized Advantage Functions (NAF) [6] address overestimation bias and streamline action selection

within continuous action spaces. Model-Based Policy Optimization (MBPO) represents model-based RL [9], while Gaussian Process Model Predictive Control (GP-MPC) blends model-based RL with Gaussian Process models for synthetic data generation and predictive control, suitable for uncertain dynamics [8,11]. Although more sample-efficient algorithms offer advantages, they can also be more unstable and complex to tune, limiting real-world applicability. Model-Agnostic Meta-Learning (MAML) excels in quickly adapting to new tasks using minimal data through initial pre-training in simulations followed by precise fine-tuning for specific real-world scenarios [2]. This method optimally utilizes the capabilities of data-intensive algorithms such as TRPO, rendering them suitable for real-world applications where data availability is constrained.

2.1 The Physical Set-Up

The AWAKE electron source and line, distinguished by its rapid repetition rate, diverse beam instrumentation, and low risk of damage due to beam loss, provides an ideal setting for the evaluation of various algorithms. The initial RL agents, aimed at trajectory optimization in the AWAKE electron line, were developed to parallel the correction efficiency of the traditional singular value decomposition (SVD) algorithm used in control rooms. These algorithms aim to guide the beam along a desired path, achieving specific parameters at the line's end required for subsequent processes. The AWAKE electron production process involves a 5MV RF gun, which accelerates the electrons to 18MeV. These electrons then traverse a 12m beam line, leading to the plasma cell. During this journey, the electron beam undergoes a vertical shift of 1 m and a 60-degree bend to align with the proton beam at the entrance of the plasma cell. The path of the electron beam is controlled using 10 horizontal and 10 vertical steering dipoles, guided by data from 10 beam position monitors (BPMs) in each plane.

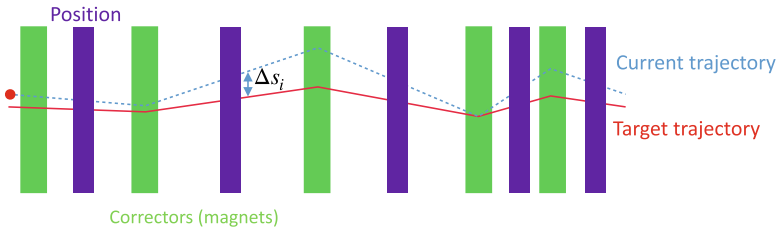


Fig. 2. Illustration of a beam steering problem as in the AWAKE electron line. Correctors (green) are followed by a Beam Position Monitor (BPM) (violet). The distance to the target is the state vector $\boldsymbol{s}$ consisting of i components, here denoted as Δs_i, where i is the index of the BPM. The measured trajectory is dashed blue and the target trajectory is red.

2.2 Simulating the Electron Transfer Line Simulations and Defining the Markov Decision Process (MDP)

The electron transfer line and its components are simulated in MAD-X [13]. This includes the transfer functions from field to current for different magnets, operating under normalized strength. Steering dipoles typically adjust trajectories by about 1mrad per corrector. Using MAD-X, the response matrix, indicative of BPM changes relative to corrector adjustments, is computed.

For maximum benefit from the simulated environment, the RL agent's observations $\boldsymbol{s}$ (BPM deviations from a reference trajectory as depicted in Fig. 2) and actions $\boldsymbol{a}$ (increments to the current dipole settings) must match the units and normalization used in the environment linked to actual equipment.

Design Decisions and MDP Details. In the live setup interfacing with AWAKE's equipment, certain measurement anomalies, like zero readings resulting from lost beam trajectories, were addressed. These readings, which fall within the anticipated observation range, were substituted with a predefined value equal to the line's aperture at the BPM location. Observations in real-machine experiments reveal moment-to-moment momentum fluctuations from the electron source, which can be easily replicated in simulations by introducing standard Gaussian noise; it is not factored into the subsequent analysis.

The reward metric is formulated as the negative root mean square (RMS) of the deviation from the target trajectory $r \propto -||\boldsymbol{s}||$ as illustrated in Fig. 2.

To add to the complexity and difficulty of the control task, initial trajectories are deliberately set far from the desired paths, and actions are restricted to maximum amplitude. This design ensures that resolving an episode is not a trivial one-step process but instead requires a more sophisticated and prolonged control strategy. The initial distribution of these trajectories is represented by ρ_0, which is illustrated in Fig. 3.

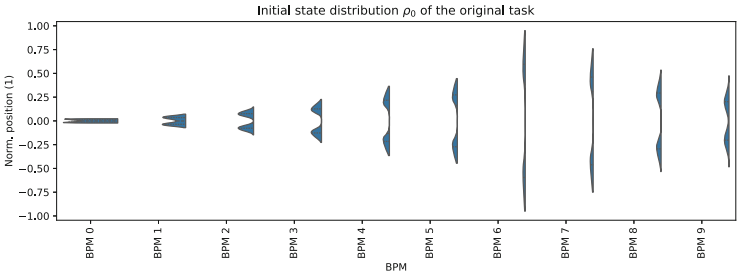


Fig. 3. Example initial distribution ρ_0 of the AWAKE MDP of task 2.

As previously noted, the precision of the position measurement apparatus restricts its capability to accurately gauge an RMS value below 1 mm. When a measurement falls beneath this specified boundary, it is inferred that the target

has been attained, prompting a reset of the episode to the initial states $\boldsymbol{s}^{\text{init}}$ sampled from ρ_0.

In instances where trajectory deviations become excessively large $||\mathbf{s}||_{\max} >=$ 10 mm, resulting in contact with the beam pipe, which possesses a diameter of 20mm, the episode undergoes a reset. Subsequently, the point of impact on the wall and all ensuing measurements are assigned a value of 10mm. This assignment is justified by the fact that the beam is effectively considered lost beyond the point of impact. An episode is also reset after a maximum number of 1000 interactions with the environment to enforce explorations.

2.3 Meta RL Approaches

Meta RL propels machine learning forward by crafting algorithms capable of swift adaptation to new tasks. Notable approaches include: MAML: Aims to find an optimal model initialization that can be quickly adapted to a wide range of tasks with minimal adjustments, leveraging gradient-based optimization to pinpoint effective parameters for fast adaptation [3]. REPTILE: Streamlines the meta-learning process by eliminating the need for complex optimization steps, thereby enhancing computational efficiency [15]. Probabilistic Model-Based Meta-Reinforcement Learning: Employs probabilistic models to forecast the outcomes of actions, facilitating rapid adaptation to new situations [16].

We concentrate on MAML for its broad applicability and effectiveness in spanning across different tasks. MAML facilitates swift task or environmental adaptation with minimal fine-tuning, enabling agents to quickly master new challenges. In this framework, RL tasks are considered as MDPs, with each task variation reflecting in the initial states, dynamics, and rewards. MAML's flexibility allows it to be tailored to a wide range of problem types, optimizing learning by adjusting the initial model parameters. This results in superior performance and adaptability compared to traditional pre-training approaches. Our application of MAML in RL incorporates advanced techniques such as action-dependent baselines and trust region methods, enhancing both the efficiency and stability of the learning process [19].

2.4 Distribution of MDP in the AWAKE Environment

In our study within the AWAKE project, we employ MAML to train agents across a varied distribution of MDPs, tailoring the system's parameters to simulate real-world dynamics. We uniformly sample the quadrupoles within the AWAKE setup by ±25% from standard settings to mimic the variability and uncertainties of the accelerator's environment, thus enhancing the agent's adaptability and readiness for real-world challenges. Figure 4 showcases the variety in linear response matrices of the MDPs for our experiments, featuring the "original task" as a benchmark-central in the distribution and derived from measurements. Alongside, we show five additional, varied MDP realizations to represent the task distribution's range, maintaining these constant to evaluate learning progress and algorithm effectiveness under potential real-world conditions. This

methodical approach aids in simulating the AWAKE machine's dynamics, allowing us to examine our algorithms' robustness and potential in managing intricate control tasks.

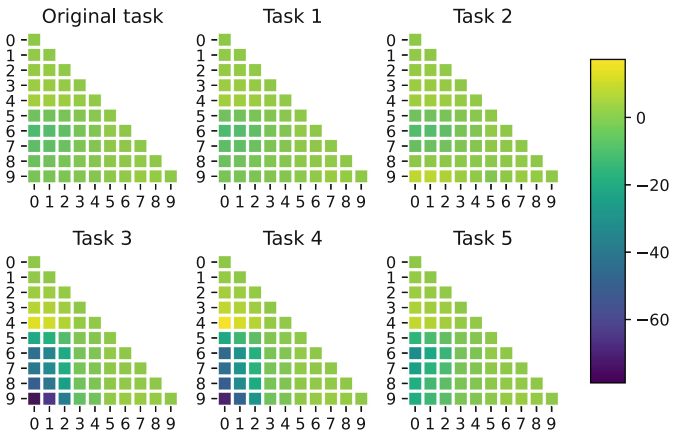


Fig. 4. The response matrices of the different test tasks as variations of the real settings. Original task corresponds to measured values of AWAKE.

3 Experiments

To assess the MAML framework's adaptability to changing MDPs, we conducted a set of focused experiments. This involved the creation of a meta-policy that encompasses a wide range of possible system changes, followed by a detailed adaptation phase in accordance with the MDP distribution framework outlined in Sect. 2.4. Throughout the training phase, a policy gradient method was utilized to fine-tune the meta-policy based on the feedback from each step of adaptation. The adaptation phase specifically targeted fixed realizations of the dynamics, employing the policy gradient methods for precise adjustments. The meta-training was done using 700,000 interactions and charting a course of continuous improvement. We evaluate three distinct initialisation scenarios for agents for the adaptation:

- **"Task 0 init."**: Utilizes a pre-trained agent, beginning with weights optimized for the central task within the distribution.
- **"Rand init."**: Initiates with randomly selected weights, providing a baseline for comparison.
- **"Meta init."**: Employs weights refined through meta-training, aiming for superior adaptability and performance.

These scenarios aim to evaluate how different initial conditions influence the performance and adaptability of agents. In Fig. 5, we examine the various training strategies, highlighting the effectiveness of the "Meta init." strategy for rapid and stable performance improvements. The "Rand init." strategy exhibits significant fluctuations in training and yields the least favorable outcomes. In contrast, "Task 0 init," demonstrates better, albeit still variable, performance. These observations underscore the potential of Meta-RL to significantly enhance adaptability and efficiency in learning within settings like the AWAKE environment. The analysis averages the results across the six predefined test tasks, as depicted in Fig. 4. This behaviour was also confirmed during a measurement campaign on the actual machine.

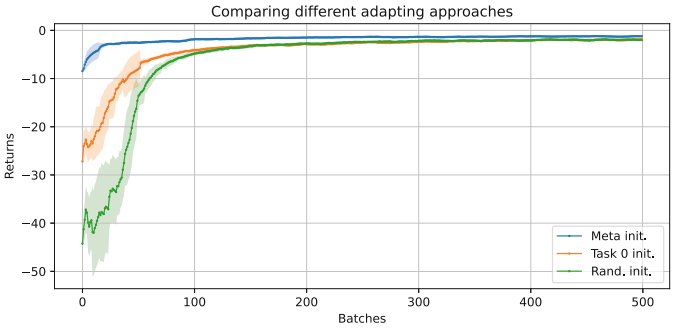


Fig. 5. A comparative evaluation of adaptation strategies, underscoring the efficacy and efficiency of the meta-training approach.

4 Conclusion

Our Meta RL research within the AWAKE project demonstrates its significant promise in enhancing adaptability and efficiency in complex systems like particle accelerators. By effectively bridging simulation and reality, Meta RL enables rapid adjustments to environmental changes and shifts in unobserved parameters, which is essential for optimal performance in dynamic settings. Despite the promise of Meta RL, particularly with MAML, challenges like computational demand and the need for detailed task analysis must be overcome. This study showcases Meta RL's capability to overcome the gap between simulation and real system in such scenarios, emphasizing the transition from theory to practical, scalable solutions. Future advances will require a collaborative approach, integrating sophisticated algorithmic methods, statistical techniques, and domain expertise to enhance system control and optimization.

References

1. Abdolmaleki, A., Springenberg, J., Tassa, Y., Munos, R., Heess, N., Riedmiller, M.: Maximum a posteriori policy optimisation. In: International Conference on Learning Representations (2018)
2. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning, Vol. 70, pp. 1126–1135 (2017)
3. Frans, K., Ho, J., Chen, X., Abbeel, P., Schulman, J.: Meta Learning Shared Hierarchies. arXiv:1710.09767 (2017)
4. Fujimoto, S., Hoof, H., Meger, D.: Addressing function approximation error in actor-critic methods. In: International Conference on Machine Learning, pp. 1587–1596 (2018)
5. Golub, G., Van Loan, C.: Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press (2013)
6. Gu, S., Lillicrap, T., Sutskever, I., Levine, S.: Continuous deep q-learning with model-based acceleration. In: International Conference on Machine Learning, pp. 2829–2838 (2016)
7. Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., et al.: Soft Actor-Critic Algorithms and Applications. arXiv:1812.05905 (2018)
8. Hirlaender, S., Lamminger, L., Zevi Della Porta, G., Kain, V.: Ultra fast reinforcement learning demonstrated at CERN AWAKE. In: Proceedings of 14th International Particle Accelerator Conference, pp. 4510–4513. JACoW Publishing, Geneva, Switzerland (2023)
9. Janner, M., Fu, J., Zhang, M., Levine, S.: When to trust your model: model-based policy optimization. In: NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 12519–12530 (2019)
10. Kain, V., Hirlander, S., Goddard, B., Velotti, F.M., Zevi Della Porta, G., Bruchon, N., Valentino, G.: Sample-efficient reinforcement learning for CERN accelerator control. Phys. Rev. Accel. Beams **23**(12), 124801 (2020). https://doi.org/10.1103/PhysRevAccelBeams.23.124801
11. Kamthe, S., Deisenroth, M.: Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control. arXiv:1706.06491 (2018)
12. Lamminger, L.: Model Based Reinforcement Learning and Meta Reinforcement Learning for Accelerator Control at CERN. Masters Thesis, PLUS University Salzburg (2023)
13. MAD-X Documentation and Source Code. Available at: https://mad.web.cern.ch/mad/
14. Michalik, D.: A Model-Based Optimal Control Approach for CERN's AWAKE Electron Line Trajectory Correction Problem. Master's Thesis, Aalborg University (2021)
15. Nichol, A., Schulman, J.: Reptile: A Scalable Metalearning Algorithm. arXiv: Learning (2018)
16. Rakelly, K., Zhou, A., Finn, C., Levine, S., Quillen, D.: Efficient off-policy meta-reinforcement learning via probabilistic context variables. In: Proceedings of the 36th International Conference on Machine Learning, Vol. 97, pp. 5331–5340 (2019)
17. Schenk, M., Combarro, E.F., Grossi, M., Kain, V., Li, K.S.B., Popa, M.-M., Vallecorsa, S.: Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines. Quantum Sci. Technol. **9**(2), 025012 (2024)

18. Schulman, J., Levine, S., Moritz, P., Jordan, M., Abbeel, P.: Trust region policy optimization. In: International Conference on Machine Learning, pp. 1889–1897 (2015)
19. Schulman, J., Moritz, P., Levine, S., Jordan, M., Abbeel, P.: High-Dimensional Continuous Control Using Generalized Advantage Estimation. arXiv:1506.02438 (2018)
20. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal Policy Optimization Algorithms. arXiv:1707.06347 (2017)

Aggregation of Inhomogeneous Non-stationary Time Series Using P-Boxes

Olgierd Hryniewicz(✉) and Katarzyna Kaczmarek-Majer

Systems Research Institute, Newelska 6, 01-447 Warszawa, Poland
{hryniewi,kkaczmarek}@ibspan.waw.pl

Abstract. Monitoring of inhomogeneous non-stationary processes has been considered. The application of well-known statistical methods for the analysis of such processes may be questionable, and in the case of long streams of data even infeasible. In the paper, we consider processes consisting of segments and subsegments. The data from subsegments belonging to respective segments are represented by histograms. For consecutive segments, they are aggregated using probability boxes (p-boxes) and a simple probabilistic method. As a result of this aggregation, consecutive segments of the monitored process are represented by triangular fuzzy numbers. These fuzzy numbers may be used for process monitoring using statistical process control (SPC) methods, such as, e.g., control charts, for fuzzy data.

1 Introduction

Real-life streams of data are usually inhomogeneous and non-stationary. This means that for different periods of observation, the collected data may be described by different probability distributions, and the parameters of these distributions may vary in time. The subject of this paper is motivated by the analysis of a real-life case - monitoring of voice characteristics recorded for patients suffering from Bipolar Disorder - a severe mental illness. In the analysis of such data, psychiatrists are interested in the monitoring patient's state of health. The observed process data come from the analysis of recorded smartphone calls. The results of observations collected during the same day should be aggregated in to evaluate the patient's state of health for this day. This problem motivated Hryniewicz and Kaczmarek-Majer [6] to aggregate subsegment data (i.e., call data) for each segment (i.e., each day), and to consider the process consisted of aggregated data.

When we need to monitor a complex segmented process, we usually need to describe it using some representative values, e.g., averages, medians, etc. Hryniewicz and Kaczmarek-Majer [6,7] described the aggregated segment data

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 184–191, 2024.
https://doi.org/10.1007/978-3-031-65993-5_22

using possibility distributions (fuzzy numbers). Then, they calculated representative values of these fuzzy numbers and used these values for the monitoring of the considered process using control charts. This approach, from a theoretical point of view, has one weak point. Possibility distributions and their characteristics (COG, alpha-cuts, etc.) do not have an established probabilistic nature, as control charts do. Therefore, the applicability of this approach can be only confirmed by simulation methods, as it was done in Hryniewicz and Kaczmarek-Majer [7]. To cope with this problem, in this paper, we propose to use another method of data aggregation, based on the probabilistic concept of p-boxes.

The paper is organized as follows. In Sect. 2, we briefly present the mathematical model of the considered segmented process. Then, in Sect. 3, we describe the proposed method of data aggregation. We use this method in Sect. 4 of the paper, where we propose a procedure that is easy to implement and can be used for the monitoring of process stability, understood as the stability of its average value. The paper is concluded in Sect. 5.

2 Mathematical Model of a Segmented Process

Let us assume that we observe a stream of numerical data consisting of a large number of observations. We assume that this stream is divided into segments, and each of these segments is further divided into subsegments. The number of subsegments in a segment, and the number of observations in a subsegment (i.e. its length) may be different in consecutive segments. The mathematical model of such a stream of data was proposed in Hryniewicz and Kaczmarek-Majer [6] and will be adopted in this paper. To make our description more clear, we will describe only one segment. We will assume that for other segments the data structure is similar. We assume that the structure of segmentation may be either known or determined unequivocally using some independent statistical or machine learning methods. In each of considered subsegments, the observed data are described by a joint probability distribution $F_j(x_{j,1}, \ldots, x_{j,n_j}), j = 1, \ldots, n$, where n_j denotes the number of observations in the jth subsegment, and n is the number of subsegments in the considered segment. These distributions are unknown but may be consistently estimated from data. Now, for purposes of the proposed data aggregation, we assume that within a given segment all subsegment data are classified into one of m classes, defined by the set of class limits $(c_0, c_1, \ldots, c_m)$. Thus, we observe discrete data $k_{j,l}, j = 1, \ldots, n, l = 1, \ldots, m$, where $k_{j,l}$ is the number of observations in the j-th subsegment that belong to the l-th class. Let $k_j = k_{j,1} + \cdots + k_{j,m}$ be the the number of observations in the j-th subsegment. Hence, the observed segment data are described by n multinomial distributions with the parameters estimated as $\hat{p}_{j,l} = k_{j,l}/k_j, j = 1, \ldots, n, l = 1, \ldots, m$.

The multinomial data described above can be considered as grouped data generated by the continuous probability distributions describing subsegment observations. If we know the functional form of these distributions, it is in principle possible to estimate the parameters of the cdf functions $F_j(x)$ describing the

subsegment data. However, for complex and inhomogeneous streams of data, the functional forms of these functions are hardly known. However, we can estimate their values only for the known class limits, using the formulae: $\hat{F}_j(c_{j,0}) = 0$ and

$$\hat{F}_j(c_{j,l}) = \sum_{l_r=1}^{l} p_{j,l_r}, j = 1, \ldots, n, l = 1, \ldots, m. \tag{1}$$

Knowing the estimated values of an unknown probability distribution only at certain discrete points, we may approximate this distribution using a piecewise linear function. This approximation is equivalent to the assumption, that unobserved data within each class are uniformly distributed, and the observed data are represented by the empirical probability density function (epdf) $\hat{p}_{j,l}, j = 1, \ldots, n, l = 1, \ldots, m$, i.e., by the respective histograms.

In many practical applications, the segmented stream of data is interpretable on the level of segments. Therefore, there is a need to aggregate all data coming from subsegments belonging to the same segment into one probability distribution describing the whole segment. Hryniewicz and Kaczmarek-Majer [6] considered two approaches to this aggregation problem. First, they proposed a simple probabilistic method leading to the construction of a probability distribution that describes a given segment. They proposed to combine subsegment epdf's by the calculation of their weighted sum with the weights proportional to the relative sizes of their respective data sets. Simple arithmetics shows that this approach is equivalent to building for each segment a combined histogram $(n_1^0, \ldots, n_m^0)$, such that $n_l^0 = k_{1,l} + \cdots + k_{n,l}$. Using the assumption that the unobserved data are uniformly distributed between class limits, this combined histogram can be easily transformed into a combined epdf, and thus into a combined cdf. Second, they proposed a possibilistic method that describes all segment data using a possibility distribution defined by a convex and normalized upper envelope of all histograms belonging to the considered segment. For such possibility distributions, they computed certain characteristic values, and used these values for process monitoring.

3 Aggregation of Segmented Processes Using p-Boxes

Destercke et al. [1] in their two papers consider mutual relationships between different models of uncertainty. According to them, dealing with the most general models presents serious challenges. Therefore, less general and simpler models, such as possibilities, see Dubois and Prade [2], or p-boxes introduced in Ferson et al. [3], are used in practice. However, it is worth noticing that the possibilities are not generalizations of the classical probability, and thus using them in purely probabilistic models for some researchers is questionable.

The concept of p-boxes was introduced by Ferson et al. [3] and was further generalized by Destercke et al. [1]. In Ferson et al. [3,4], the authors list the serious practical problems that may be solved using this concept, including the problem of non-stationarity (non-constant distributions) and interval-censoring.

Taking into account the data model considered in this paper and the representation of measured data by histograms, we can assume that p-boxes may be considered the appropriate model for the analysis of this type of data.

Ferson et al. [3] defined the p-box in the following way. Suppose $\overline{F}$ and $\underline{F}$ are nondecreasing functions from the real line $\mathcal{R}$ to $[0,1]$ and $\underline{F} \leq \overline{F}$ for all $x \in \mathcal{R}$. Let $[\underline{F}, \overline{F}]$ denote the set of all nondecreasing functions F from $\mathcal{R}$ into $[0,1]$ such that $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$. When the functions $\underline{F}$ and $\overline{F}$ circumscribe an imprecisely known probability distribution, we call $[\overline{F}, \underline{F}]$, specified by this pair of functions, a "probability box" or a "p-box" for that distribution. This means that, if $[\overline{F}, \underline{F}]$ is a p-box for a random variable X whose distribution F is unknown except that it is within the p-box, then $\underline{F}(x)$ is a lower bound on $F(x)$ which is the (imprecisely known) probability that the random variable X is smaller than x. Likewise, $\overline{F}(x)$ is an upper bound on the same probability. Destercke et al. [1] presented an equivalent definition of a p-box using a general concept of credal sets. Then, they generalized the classical concept of a p-box by assuming a more general space X (classical p-boxes are defined on a subset of the real line). This generalization lets them analyze mutual relationships between p-boxes and other representations of imprecise information like random sets, probability intervals, and possibilities.

Now, let us consider the aggregation of grouped data from n subsegments that belong to the same segment. Then, the estimated cdf's for grouped data, calculated according to (1) may be denoted as $(\hat{F}_j(c_0), \ldots, \hat{F}_j(c_m)), j = 1, \ldots, n$. Using these values we can estimate the upper and lower bounds of the p-box that describes the data belonging to the considered segment

$$\overline{\hat{F}}(c_l) = \max_j \hat{F}_j(c_l), l = 0, \ldots, m \tag{2}$$

$$\underline{\hat{F}}(c_l) = \min_j \hat{F}_j(c_l), l = 0, \ldots, m \tag{3}$$

Let us illustrate this method of aggregation with a real-life example of a voice characteristic *zero crossing rate*, recorded for 5 smartphone calls (subsegments) of a BD patient during one day (segment). The range of possible values of this characteristic ($[0,1]$) was divided into 10 equidistant classes, and the results of measurements (restricted to the first 8 classes) were given in a grouped form in Table 1.

Table 1. Measurement data of the *zcr* voice characteristic

Call-1	1260	794	114	38	16	18	16	8
Call-2	525	558	117	30	20	16	2	0
Call-3	25	390	1021	1308	964	748	420	48
Call-4	7	141	448	412	282	232	145	12
Call-5	872	602	85	26	24	11	1	0

It is easy to notice that the recorded calls are of different lengths, and the probability distributions of 2 calls are significantly different than the probability distributions of the remaining ones.

When we look for real-life applications of p-boxes we usually need to calculate expectations of a random variable X characterized by the p-box $(\overline{F}(x), \underline{F}(x))$ with respect to a certain real-valued function $h(x)$. In general, this is a very difficult task, but it can be solved relatively easily using some approximations. Utkin and Destercke [9] approximate the distribution function $F(x)$ by the set of their N values $F(x_i), i = 1, \ldots, N$. Using this approximation, they find the approximate upper (lower) expectation of the random variable X by solving two linear programming problems (one for maximization and one for minimization). A simpler solution can be found if $h(x)$ is a monotone non-decreasing (non-increasing) function. In this case, the upper and lower expectations are completely determined by $\overline{F}(x)$ and $\underline{F}(x)$ functions. However, the functional form of these functions is usually unknown, and even in this special case, we have to use their approximations.

When $h(x) = x$, i.e. when we want to calculate the classical expected value of X, and our data are described by $\overline{F}(x)$ and $\underline{F}(x)$ functions, the lower and upper expectations are calculated from the formulae (see Ferson et al. [4] for a more general case of interval-valued data).

$$\overline{E} = \sum_{i=1}^{m} c_i * (\underline{F}(c_i) - \underline{F}(c_{i-1}) \ , \underline{E} = \sum_{i=1}^{m} c_{i-1} * (\overline{F}(c_i) - \overline{F}(c_{i-1}). \qquad (4)$$

The calculation of the lower and upper variances is not so simple, but the respective algorithm can be found in Ferson et al. [4].

4 Monitoring of Segmented Processes Using Their Representations by Fuzzy Triangular Numbers

Control charts are simple tools for the monitoring of process stability using representative values (such as average, median, variance, etc.) that have probabilistic interpretation. In contrast to the possibilistic approach proposed in Hryniewicz and Kaczmarek-Majer [7], the proposed method of aggregation, based on the concept of p-boxes, is purely probabilistic. Thus, one can build control charts for monitoring process characteristics that represent probability distributions $\overline{F}(x)$ and $\underline{F}(x)$.

However, from a practical point of view, this approach may be questionable. First of all, the p-box limits, $\overline{F}(x)$ and $\underline{F}(x)$, as it is a usual case for the results of min-max optimization, give quite a conservative representation of observed data. Look, for example, at their expected values, calculated according to (4). The values of $\underline{E}$ ($\overline{E}$) are calculated in such a way, that all process data have been concentrated at lower (upper) class limits. This situation is highly unrealistic, and cannot be recommended for building control charts. Therefore, we propose a different method for the construction of approximate p-box expectations for grouped data. We can assume that instead of discrete-valued cdfs we have the

following piecewise-linear cdfs

$$\overline{\hat{F}}^{\star}(x) = \overline{\hat{F}}(c_{l-1}) + \frac{x - \overline{\hat{F}}(c_{l-1})}{\overline{\hat{F}}(c_l) - \overline{\hat{F}}(c_{l-1})}, l = 1, \ldots, m, \tag{5}$$

$$\underline{\hat{F}}^{\star}(x) = \underline{\hat{F}}^{\star}(c_{l-1}) + \frac{x - \underline{\hat{F}}(c_{l-1})}{\underline{\hat{F}}(c_l) - \underline{\hat{F}}(c_{l-1})}, l = 0, \ldots, m, \tag{6}$$

where functions $\overline{\hat{F}}(x)$ and $\underline{\hat{F}}(x)$ are given by (2)–(3), respectively. Note, that this proposal is based on the assumption that the random variables represented, respectively, by the cdfs $\overline{\hat{F}}(x)$ and $\underline{\hat{F}}(x)$ are uniformly distributed between the consecutive class limits $c_0, c_1, \ldots, c_m$. This assumption is often used in descriptive statistics for the analysis of grouped data represented by histograms.

The expected values of random variables distributed according to (5)–(6) can be calculated as follows

$$\overline{E}^{\star} = \sum_{i=1}^{m} \frac{c_{i-1}+c_i}{2} * (\underline{\hat{F}}(c_i) - \underline{\hat{F}}(c_{i-1})) \ , \ \underline{E}^{\star} = \sum_{i=1}^{m} \frac{c_{i-1}+c_i}{2} * (\overline{\hat{F}}(c_i) - \overline{\hat{F}}(c_{i-1})). \tag{7}$$

In Statistical Process Control (SPC), control charts are used to generate alarm signals that the probability distribution describing process observations has been changed. In process monitoring, we usually monitor either central tendency (the so-called 'process level') or variability. An alarm signal is generated when the value of a monitored statistic (e.g., an average), plotted on a control chart, falls beyond a certain control limit. The p-boxes, considered in this paper, are rather focused on the description of borderline extremes, and not on their impact on a process level. Therefore, we need a variable that can provide additional information about a certain '*average*' of the process level. For this purpose, we can use the probabilistic aggregation method introduced by Hryniewicz and Kaczmarek-Majer [6] and described in Sect. 2. Let $n_1^0, \ldots, n_m^0$ denote a combined histogram for a considered segment with the total number of N observations. Thus, we can calculate for these cumulated data the average and its lower and upper bounds

$$\underline{E}_C^{\star} = \sum_{i=1}^{m} c_{i-1} \frac{n_i}{N} \ , \ E_C^{\star} = \sum_{i=1}^{m} \frac{c_{i-1}+c_i}{2} \frac{n_i}{N} \ , \ \overline{E}_C^{\star} = \sum_{i=1}^{m} c_i \frac{n_i}{N} \tag{8}$$

The bounds $\underline{E}_C^{\star}$ and $\overline{E}_C^{\star}$ may not reflect the variety of subsegment data distributions. Therefore, we propose to describe each segment of the considered process by the triple $(\underline{E}^{\star}, E_C^{\star}, \overline{E}^{\star})$. Such triples, calculated for consecutive segments, can be plotted on a control chart. For the calculation of control limits of the chart, we can use the values of $E_C^{\star}$ computed for consecutive process segments, according to an algorithm depending on the type of a chart (see Montgomery [8] for information). An example of such a chart, calculated from real-life data (voice *zcr* characteristic observed for a BD patient) is shown in Fig. 1.

In the considered case, the monitored process is described by three statistics, and a question arises then, how to generate alarm signals. We propose to use

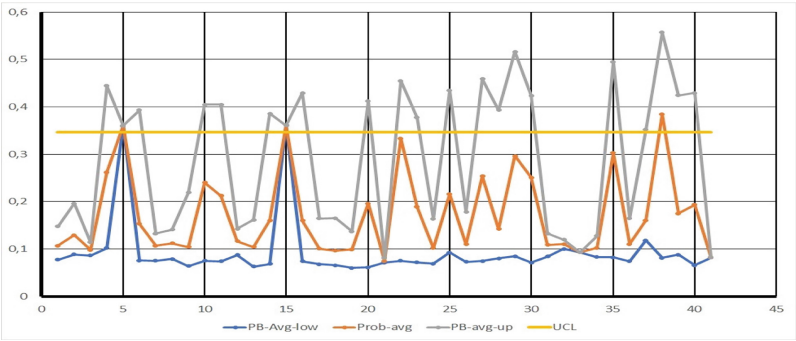


Fig. 1. BD calls data - expected p-box values of the voice *zcr* characteristic.

the triple $(\underline{E}^\star, E_C^\star, \overline{E}^\star)$ as the description of a fuzzy triangular number $\tilde{T}$, and to use such fuzzy numbers for making decisions using a control chart that operates according to the rules proposed by Grzegorzewski and Hryniewicz [5]. Let C_U (C_L) be an upper (lower) control limit and $\alpha \in (0, 1]$ be a certain critical value. The alarm is generated when a certain measure of the strength of the relation $\tilde{T} > C_U$ or $\tilde{T} < C_L$, e.g., its *necessity* (see Dubois and Prade [2]), is not smaller than α. In the considered case, the respective *necessity* measure can be calculated from the formulae

$$Nec(\tilde{T} > C_U) = \begin{cases} 0 & , E_C^\star < C_U \\ \frac{C_U - E_C^\star}{E_L^\star - E_C^\star} & , \text{otherwise} \\ 1 & , E_U^\star > C_U \end{cases} \quad Nec(\tilde{T} < C_L) = \begin{cases} 0 & , E_C^\star > C_L \\ \frac{E_U^L - C^L}{E_U^\star - E_C^\star} & , \text{otherwise} \\ 1 & , E_U^\star < C_L \end{cases} \tag{9}$$

One can also consider a weaker criterion for the generation of alarms (not described in this paper) based on the concept of *possibility* (see Dubois and Prade [2]). If we look at data plotted in Fig. 1 we can see that for segments 1 and 15 the necessity that $\tilde{T} > C_U$ is equal to one (each of these segments consists of only one subsegment). This necessity is also positive (0.22) for segment 38.

5 Conclusions

In the paper we described a segmented non-stationary and complex process using the concept of p-boxes. The p-boxes were used for an imprecise description of aggregated segment data. The two expected values (lower and upper) of the p-box that describe a segment of data, together with the expected value of its probabilistically aggregated distribution, describe the segment's average by a triangular fuzzy number. These fuzzy numbers are plotted on a control chart that is used for the monitoring of process stability. Alarm signals on this chart are generated when the necessity that the triangular fuzzy number is greater

(smaller) than a given upper (lower) control limit is greater than a certain number α, chosen from the interval $(0, 1]$. Using this approach we can monitor the stability of process measures of central tendency. The usage of this approach for monitoring process variability needs further research. The proposed approach is not computationally demanding, so it can be also used for monitoring continuous streams of data measured at discrete points.

References

1. Destercke, S., Dubois, D., Chojnacki, E.: Unifying practical uncertainty propagations: I generalized p-boxes, II clouds. Int. J. Approx. Reason. **49**, 649–677 (2008)
2. Dubois, D., Prade, H.: Possibility Theory—An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
3. Ferson, F., Ginzburg, L., Kreinovich, V., Myers, D., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015. Sandia National Laboratories (2003)
4. Ferson, F., Kreinovich, V., Hajagos, J., Oberkampf, W., Ginzburg, L.: Experimental uncertainty estimation and statistics for data having interval uncertainty. Technical Report SAND2007-0939. Sandia National Laboratories (2007)
5. Grzegorzewski, P., Hryniewicz, O.: Soft methods in statistical quality control. Control. Cybern. **29**(1), 1–22 (2000)
6. Hryniewicz, O., Kaczmarek-Majer, K.: Possibilistic aggregation of inhomogeneous streams of data. In: Proceedings of FUZ-IEEE Conference (2021). https://doi.org/10.1109/FUZZ45933.2021.9494583
7. Hryniewicz, O., Kaczmarek-Majer, K.: Monitoring of possibilistic aggregated complex time series. In: Garcia-Escudero, L.A. et al. (Eds.) Building Bridges Between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing, Vol. 1433, pp. 208–215. Springer (2023)
8. Montgomery, D.C.: Introduction To Statistical Quality Control, 6th edn. Wiley, New York (2011)
9. Utkin, L., Destercke, S.: Computing expectations with continuous p-boxes: univariate case. Int. J. Approx. Reason. **50**, 778–798 (2009)

Entropy-Based Search for the Most Informative Belief Functions

Radim Jiroušek and Václav Kratochvíl(✉)

Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic
{radim,velorex}@utia.cas.cz

Abstract. The paper deals with the problem studied in our previous paper published in Int. J. Approx. Reasoning, which raised new questions rather than brought solutions. Thus, the current contribution also tries to answer the ever-lasting question: Which belief function entropies described in the literature can detect optimal models? Nevertheless, here, we approach the problem differently. We try to find out the entropy functions that are indirectly proportional to the informative content of belief functions, i.e., the moreinformative the belief function, the lower its entropy.

Keywords: belief functions · entropy · information content

1 Introduction

In probability theory, many data-based model learning approaches use the Shannon entropy of a model as a criterion for optimality. Shannon entropy quantifies the uncertainty in a probability distribution, meaning that the higher entropy, the less information is carried by the distribution. Therefore, when choosing the optimal model from a set that includes all necessary information, one should apply the maximum entropy principle. This principle ensures adding the least amount of "artificial" information.

Conversely, the minimum entropy principle is appropriate when selecting the optimal model from those that contain only the necessary information. This principle favors models that incorporate as much required information as possible. In this paper, we explore the measurement of information content of belief functions and seek to identify metrics helping to discern which of the two belief functions is more informative. We do not, however, differentiate between types of uncertainty assessed by these measures (like nonspecificity, dissonance, confusion, and others).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 192–199, 2024.
https://doi.org/10.1007/978-3-031-65993-5_23

2 Notation

We will get along with only a few fundamental concepts from the belief function theory. Let Ω denote a finite set, often called a *frame of discernment*. A *basic probability assignment* (BPA) is a function $m : 2^{\Omega} \to [0, 1]$ satisfying two conditions: (1) $\sum_{a \subseteq \Omega} m(a) = 1$, and (2) $m(\emptyset) = 0$.

A subset $a \subseteq \Omega$ is said to be a *focal element* of m if $m(a) > 0$. A BPA with only one focal element is said to be *deterministic*, denoted ι_a, where $\iota_a(a) = 1$. As ι_Ω represents total ignorance, it is termed *vacuous*. A BPA is called *Bayesian* if all its focal elements are singletons.

Each BPA m is associated with the following three functions defined over the power set of the frame of discernment:

$$Bel_m(a) = \sum_{b \subseteq a} m(b); \quad Pl_m(a) = \sum_{b \subseteq \Omega : b \cap a \neq \emptyset} m(b); \quad Q_m(a) = \sum_{b \subseteq \Omega : b \supseteq a} m(b).$$

These are known respectively as the *belief function, plausibility function*, and *commonality function*.

Each BPA m is also connected with a set of probability distributions. It is called *credal set*, and it is defined:

$$\mathcal{P}_m = \left\{ Pr \text{ defined on } \Omega : \left(\forall a \subseteq \Omega : Pr(a) \geq Bel(a) \right) \right\}.$$

A central concept in Dempster-Shafer's theory is Dempster's combination rule [17], which combines information from two distinct sources: BPAs m_1 and m_2. The combined BPA $m_1 \oplus m_2$ is calculated (for each subset $c \subseteq \Omega$) as follows:

$$m_1 \oplus m_2(c) = (1 - K)^{-1} \sum_{a \subseteq \Omega} \sum_{b \subseteq \Omega : a \cap b = c} m_1(a) \cdot m_2(b),$$

where $K = \sum_{a \subseteq \Omega} \sum_{b \subseteq \Omega : a \cap b = \emptyset} m_1(a) \cdot m_2(b)$ is often interpreted as the amount of conflict between m_1 and m_2 (if $K = 1$, then the combination is undefined).

In definitions of entropy functions for BPAs (see Table 1 for the list of those considered in this paper), some authors also used Shannon entropy of a specific probability distribution (recall Shannon entropy is defined for probability function Pr by the formula $\mathcal{H}(Pr) = -\sum_{\omega \in \Omega} Pr(\omega) \log_2 Pr(\omega)$). Thus, in what follows, we will consider three probability functions related to BPA m, so-called *pignistic transform*, *plausibility transform*, and *maximum entropy transform* defined (respectively).

$$Bet_P_m(\omega) = \sum_{a \subseteq \Omega : \omega \in a} \frac{m(a)}{|a|},$$

$$Pl_P_m(\omega) = \frac{Pl_m(\omega)}{\sum_{v \in \Omega} Pl_m(v)},$$

$$ME_P_m = \arg \max_{P \in \mathcal{P}_m} \{\mathcal{H}(P)\}.$$

Table 1. Definitions of entropy, chronologically ordered

H_O	Höhle [4]	$H_O(m) = \sum_{a \subseteq \Omega} m(a) * \log\left(\frac{1}{Bel_m(a)}\right)$
H_T	Smets [18]	$H_T(m) = \sum_{a \subseteq \Omega} \log\left(\frac{1}{Q_m(a)}\right)$
H_D	Dubois and Prade [2]	$H_D(m) = \sum_{a \subseteq \Omega} m(a) \log(\lvert a \rvert)$
H_N	Nguyen [13]	$H_N(m) = \sum_{a \subseteq \Omega} m(a) * \log\left(\frac{1}{m(a)}\right)$
H_L	Lamata and Moral [11]	$H_L(m) = H_Y(m) + H_D(m)$
H_K	Klir and Ramer [10]	$H_K(m) = \sum_{a \subseteq \Omega} m(a) * \log\left(\frac{1}{1-\sum_{b \subseteq \Omega} m(b) \frac{\lvert b \setminus a \rvert}{\lvert b \rvert}}\right) + H_D(m)$
H_P	Klir and Parviz [9]	$H_P(m) = \sum_{a \subseteq \Omega} m(a) * \log\left(\frac{1}{1-\sum_{b \subseteq \Omega} m(b) \frac{\lvert a \setminus b \rvert}{\lvert a \rvert}}\right) + H_D(m)$
H_B	Pal et al. [14,15]	$H_B(m) = H_D(m) + H_N(m)$
H_I	Maeda and Ichihashi [12]	$H_I(m) = H_H(m) + H_D(m) = \mathcal{H}(ME_P_m) + H_D(m)$
H_H	Harmanec and Klir [3]	$H_H(m) = \max_{P \in \mathcal{P}_m} \mathcal{H}(P) = \mathcal{H}(ME_P_m)$
H_J	Jousselme et al. [8]	$H_J(m) = \mathcal{H}(Bet_P_m)$
H_Y	Yager [19]	$H_Y(m) = \sum_{a \subseteq \Omega} m(a) * \log\left(\frac{1}{Pl_m(a)}\right)$
H_G	Deng [1]	$H_G(m) = H_N(m) + \sum_{a \subseteq \Omega} m(a) * \log(2^{\lvert a \rvert} - 1)$
H_λ	Jiroušek and Shenoy [5]	$H_\lambda(m) = \mathcal{H}(Pl_P_m) + H_D(m)$
H_S	Jiroušek and Shenoy [6]	$H_S(Q_m) = \sum_{a \subseteq \Omega} (-1)^{\lvert a \rvert} Q_m(a) \log(Q_m(a))$
H_π	Jiroušek et al. [7]	$H_\pi = \mathcal{H}(Bet_P_m) + H_D(m)$

3 Entropy Functions

The study is based on the intuition saying that BPA m_1 is not less informative than BPA m_2 (assuming that both are defined on the same frame of discernment Ω) if $Bel_{m_2} \leq Bel_{m_1}$, which is equivalent to $Pl_{m_1} \leq Pl_{m_2}$, and also to $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$. Notice that this situation is very general and covers some other specific cases. For instance, consider a situation where each focal element c of m_1 corresponds to a focal element $\bar{c}$ of m_2, such that $m_1(c) = m_2(\bar{c})$ and $c \subseteq \bar{c}$. In a way, the simplest case is the following: We say that m_1 is a *simple specification* of m_2 if m_1 is created from m_2 by shifting a part of its mass from some focal element to its subset; precisely, there exist subsets $a \subset b \subseteq \Omega$ such that $m_1(a) = m_2(a) + \varepsilon$, and $m_1(b) = m_2(b) - \varepsilon$ (all the remaining focal elements of m_1 are the copies of the focal elements of m_2). Since we shift (a part of) the mass from b to its subset, we see directly from the definition of belief function that[1] $Bel_{m_1} > Bel_{m_2}$.

In this paper, we say that an entropy H is *simply monotonic* if for every simple specification m_1 of m_2, it holds that $H(m_1) < H(m_2)$. This concept of simple monotonicity is less stringent than the notions of monotonicity discussed by [12,16]. So, from the cited papers, we immediately see that H_I and H_D are simply monotonic. On the other hand, proving that an entropy H is simply

[1] Strict inequality $Bel_{m_1} > Bel_{m_2}$ in this paper denotes that for all $a \subseteq \Omega$, $Bel_{m_1}(a) \geq Bel_{m_2}(a)$, and for at least one a $Bel_{m_1}(a)$ is strictly greater than $Bel_{m_2}(a)$.

monotonic does not guarantee that it is also monotonic in the sense used by the cited authors. To be clear, we will consistently use the notion of simple monotonicity because, in some other papers, the monotonicity is also used in a quite different meaning (see, e.g., [5]).

One could also consider another type of monotonicity related to credal sets. Nevertheless, such a type of monotonicity is equivalent to the introduced simple monotonicity since, as it is shown in the following assertions, if $Bel_{m_2} < Bel_{m_1}$, then m_1 may be created from m_2 by a sequence of simple specifications.

Theorem 1. *If $Bel_m < Bel_{\bar{m}}$, then there exists a finite sequence of BPAs $m = m_1, m_2, \ldots, m_k = \bar{m}$ such that each m_{i+1} is a simple specification of m_i.*

Proof. The assertion is a direct consequence of the following lemma, the proof of which gives instructions on how to construct m_{i+1} from m_i. Notice that, in comparison with m_i, m_{i+1} always has more focal elements identical with focal elements of $\bar{m}$, which guarantees that the constructed sequence of BPAs $m_1, m_2, \ldots, m_k$ is finite. □

Lemma 1. *If $Bel_{m_1} < Bel_{m_2}$, then there exists a BPA $\hat{m}$, which is a simple specification of m_1. Moreover*

$$|\{a \subseteq \Omega : m_1(a) = m_2(a)\}| < |\{a \subseteq \Omega : \hat{m}(a) = m_2(a)\}|.$$

Proof. Since $m_1 \neq m_2$ and both these BPAs are normalized, a focal element b of m_1 must exist such that $m_1(b) > m_2(b)$. The existence of $a \subset b$ for which $m_1(a) < m_2(a)$ follows from the following contemplation: If $m_1(a) \geq m_2(a)$ for all $a \subseteq b$, then $Bel_{m_1}(b) > Bel_{m_2}(b)$.

Choose $\varepsilon = \min\{(m_2(a) - m_1(a)); (m_1(b) - m_2(b))\}$. Define $\hat{m}(a) = m_1(a) + \varepsilon$, $\hat{m}(b) = m_1(b) - \varepsilon$, and $\hat{m}(c) = m_1(c)$ for all $c \subseteq \Omega$ different from a and b, which is a simple specification of m_1. Moreover, if $\varepsilon = m_2(a) - m_1(a)$, then $\hat{m}(a) = m_2(a)$, and if $\varepsilon = m_1(b) - m_2(b))$, then $\hat{m}(b) = m_2(b)$. Since all the remaining values of $\hat{m}$ are the same as the corresponding values of m_1, it means that

$$|\{a \subseteq \Omega : m_1(a) = m_2(a)\}| < |\{a \subseteq \Omega : \hat{m}(a) = m_2(a)\}|.$$

□

Results of Experimental Computations. Since there are no more theoretically supported properties expressing the relationship between entropy functions and the informativeness of belief functions, we made a lot of computations. Because of the high computational complexity of H_H and H_I, we exclude them from the experiments. It does not deprive our results because Harmanec-Klir's entropy H_H is, apparently, simply monotonic with reservation that the required inequality is not strict, and, as said above, the simple monotonicity of Maeda-Ichihashi's entropy H_I follows from the results in [12].[2]

[2] The simple monotonicity of $H_I = H_D + H_H$ also follows directly from the properties of H_D and H_H.

The experiments were designed as follows: Starting with a frame of discernment, Ω, with a size of $|\Omega| = 16$, we randomly generated BPAs denoted as m_2. Subsequently, for each m_2 we randomly created m_1 as a simple specification of m_2. Then, we computed all the considered entropy values for m_1 and m_2 to compare whether the entropy $H(m_1)$ was less than $H(m_2)$, indicating that m_1 contains more information. The result was positive if $H(m_1) < H(m_2)$, and negative otherwise. This computational process was repeated 2,000 times.

Our results, illustrated in Fig. 1, demonstrate the effectiveness of various entropy measures in identifying increased information content in m_1 over m_2. Notably, four entropy measures H_T, H_D, H_λ, and H_π exhibited a perfect success rate of 100%. For H_D, it is consistent with what was said above. However, we lack theoretical evidence for H_T, H_λ and H_π to assert that they are simply monotonic. Unlike H_I, which is defined as a sum of H_D and the Shannon entropy of the corresponding max-entropy transform, neither the pignistic nor the plausibility transform guarantees simple monotonicity of H_λ and H_π.

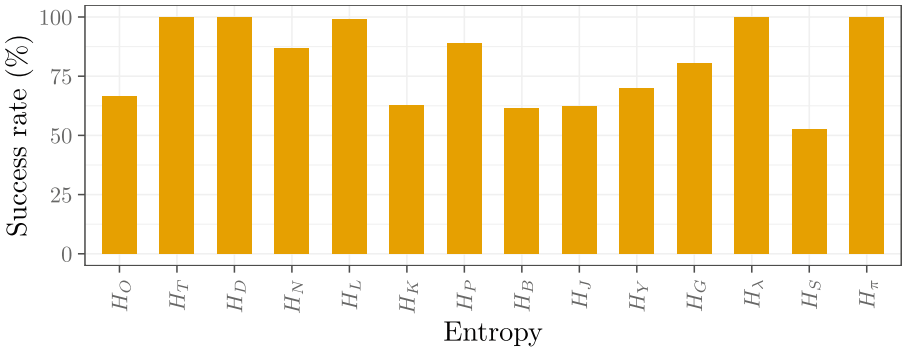


Fig. 1. Success rate simple-specification experiments.

To illustrate this point, consider a BPA m_1 defined on a binary frame $\Omega = \{0, 1\}$ with focal elements $m(0) = \frac{1}{5}$ and $m(\Omega) = \frac{4}{5}$. The Shannon entropy of both the pignistic and plausibility transforms of m_1 and its simple specifications, m_2 and m_3, are presented in Table 2. This table reveals that, among others, H_J is not simply monotonic.

Close behind the four above-mentioned entropies, H_L shows a success rate exceeding 99%. H_L is a sum of two other entropy measures ($H_D + H_Y$) where H_D is 100% successful, implying that this amalgamation suggests limited benefits from mixing these entropy measures. For a simple example where H_L fails to reflect the increase in informativeness for simple specifications, see Table 3. Note that the problem is the imbalance between the sizes of the mixed entropies H_D and H_Y. It was very simple to find an example with $|\Omega| = 2$. As $|\Omega|$ increases, the difference between the values of H_D and H_Y decreases, making it more difficult to construct such an example.

Table 2. BPA m_1 with its simple specifications m_2 and m_3.

	m_i			$Bet_P_{m_i}$			$Pl_P_{m_i}$		
	{0}	{1}	{0,1}	{0}	{1}	$\mathcal{H}(Bet_P_{m_i})$	{0}	{1}	$\mathcal{H}(Bet_P_{m_i})$
m_1	$\frac{1}{5}$	0	$\frac{4}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	0.97	$\frac{5}{9}$	$\frac{4}{9}$	0.99
m_2	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{2}$	$\frac{1}{2}$	1.00	$\frac{1}{2}$	$\frac{1}{2}$	1.00
m_3	$\frac{1}{5}$	$\frac{4}{5}$	0	$\frac{1}{5}$	$\frac{4}{5}$	0.72	$\frac{1}{5}$	$\frac{4}{5}$	0.72

Table 3. H_L for BPA m_1 and its simple specifications m_2.

	m_i			Pl_{m_i}					
	{0}	{1}	{0,1}	{0}	{1}	{0, 1}	$H_Y(m_i)$	$H_D(m_i)$	$H_L(m_i)$
m_1	0.30	0.60	0.10	0.40	0.70	1.00	0.49	0.07	0.56
m_2	0.35	0.60	0.05	0.40	0.65	1.00	0.58	0.03	0.61

4 Application of Dempster's Rule

The content of this section is built on the idea that $m_1 \oplus m_2$, the result of combining two distinct sources of information, should be more informative than each separately. In this situation, we do not have any theoretical support; we can only present.

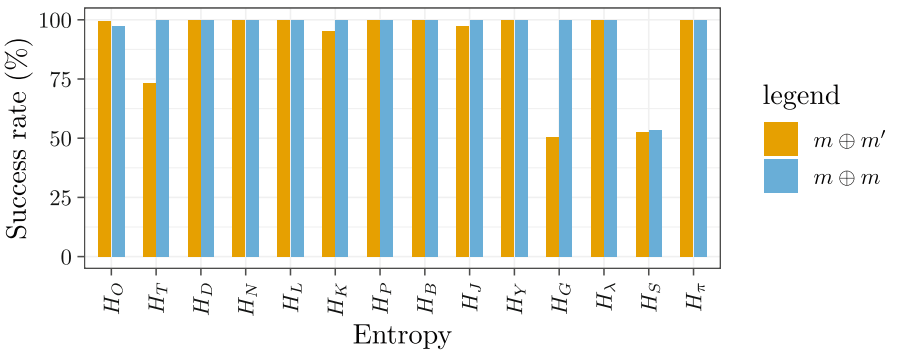


Fig. 2. Success rate for Dempster's-rule experiments.

Results of Experimental Computations. Because of the above-mentioned reasons, we again excluded both H_H and H_I from the experiments. We initiated them by generating 2,000 pairs of basic probability assignments (BPAs), labeled as m_1 and m_2 (both defined on the same frame of discernment Ω, $|\Omega| = 16$). We combined each pair using Dempster's rule of combination, getting $m_1 \oplus m_2$. The primary objective was to determine how often the entropy of the combination,

$H(m_1 \oplus m_2)$, was lower or equal to the initial entropy $H(m_2)$, i.e., how often the individual entropies indicate an increase of information as an impact of considering two sources of information instead of only one of them. In these experiments, we accept for success even when $H(m_2) = H(m_1 \oplus m_2)$. It is because, for some specific m_1, it may happen that $m_2 = m_1 \oplus m_2$.

The outcomes of these comparisons, presented in Fig. 2, revealed a striking contrast to our previous experiment involving simple specifications. Remarkably, eight entropy measures achieved a perfect success rate of 100%: H_D, H_L, H_λ, H_N, H_P, H_π, H_Y, and H_B. Surprisingly, Deng's entropy H_G failed - exhibiting behavior bordering on randomness. Similarly with H_S and H_T.

5 Discussions and Open Problems

There is little to add to the results from Sect. 3. The simple monotonicity saying that m_2 contains more information than m_1 if $Bel_{m_1} < Bel_{m_2}$ fully corresponds with our intuition. It holds for H_D (which indicates the close connection with non-specificity of BPAs as discussed by [10]) and H_I, and it is still an open question whether it is met also for H_T, H_λ and H_π.

Section 4 raises more questions, even though the results in Fig. 2 hint that most of the studied entropies manifest the expected property. Because of the lack of space, we mention only the most important objection. The basic assumption that *the result of combining two distinct sources of information is more informative than each separately* is somewhat questionable. It is clear that when combining two reliable sources of information, we know more (or at least the same) than when having just one of them. It, however, does not mean that we must be more specific about the possibilities in question. Consider, for example, two sources yielding Bayesian BPAs on $\Omega = \{1, 2, 3, 4\}$: $m_1 = (0.1, 0.1, 0.8, 0)$, and $m_2 = (0.1, 0.1, 0, 0.8)$. Then, $m_1 \oplus m_2 = (0.5, 0.5, 0, 0)$. Having Bayesian BPA, we can compute their Shannon entropies getting $\mathcal{H}(m_1) = \mathcal{H}(m_2) \doteq 0.92$ and $\mathcal{H}(m_1 \oplus m_2) = 1$. This example demonstrates that entropies enabling evaluation of the informativeness of belief functions should perhaps meet some other properties and predetermines, thus, future research.

Acknowledgements. We want to thank Prakash P. Shenoy, whose inspiring discussions motivated the research, and Erik Quaeghebeur for his bright advice at the right time. The Czech Science Foundation supported the second author by Grant No. 21-07494S.

References

1. Deng, Y.: Deng entropy. Chaos Solit. Fract. **91**, 549–553 (2016)
2. Dubois, D., Prade, H.: Properties of measures of information in evidence and possibility theories. Fuzzy Sets Syst. **24**(2), 161–182 (1987)
3. Harmanec, D., Klir, G.J.: Measuring total uncertainty in Dempster-Shafer theory: a novel approach. Int. J. Gen. Syst. **22**(4), 405–419 (1994)

4. Höhle, U.: Entropy with respect to plausibility measures. In: Proceedings of 12th IEEE International Symposium on Multiple Valued Logic. pp. 167–169. Paris (1982)
5. Jiroušek, R., Shenoy, P.P.: A new definition of entropy of belief functions in the Dempster-Shafer theory. Int. J. Approx. Reason. **92**, 49–65 (2018)
6. Jiroušek, R., Shenoy, P.P.: On properties of a new decomposable entropy of Dempster-Shafer belief functions. Int. J. Approx. Reason. **119**, 260–279 (2020)
7. Jiroušek, R., Kratochvíl, V., Shenoy, P.P.: Entropy for evaluation of Dempster-Shafer belief function models. Int. J. Approx. Reason. **151**, 164–181 (2022)
8. Jousselme, A.L., Liu, C., Grenier, D., Bossé, É.: Measuring ambiguity in the evidence theory. IEEE Trans. Syst. Man Cybern. A Syst. Hum. **36**(5), 890–903 (2006)
9. Klir, G.J., Parviz, B.: A note on the measure of discord. In: Uncertainty in Artificial Intelligence, pp. 138–141. Elsevier (1992)
10. Klir, G.J., Ramer, A.: Uncertainty in the Dempster-Shafer theory: a critical re-examination. Int. J. Gen. Syst. **18**(2), 155–166 (1990)
11. Lamata, M.T., Moral, S.: Measures of entropy in the theory of evidence. Int. J. Gen. Syst. **14**(4), 297–305 (1988)
12. Maeda, Y., Ichihashi, H.: An uncertainty measure with monotonicity under the random set inclusion. Int. J. Gen. Syst. **21**(4), 379–392 (1993)
13. Nguyen, H.T.: On entropy of random sets and possibility distributions. In: Bezdek, J. (Ed.) The Analysis of Fuzzy Information. Mathematics and Logic, Vol. 1, pp. 145–156. CRC Press (1987)
14. Pal, N.R., Bezdek, J.C., Hemasinha, R.: Uncertainty measures for evidential reasoning I: a review. Int. J. Approx. Reason. **7**(3–4), 165–183 (1992)
15. Pal, N.R., Bezdek, J.C., Hemasinha, R.: Uncertainty measures for evidential reasoning II: a new measure of total uncertainty. Int. J. Approx. Reason. **8**(1), 1–16 (1993)
16. Ramer, A.: Uniqueness of information measure in the theory of evidence. Fuzzy Sets Syst. **24**(2), 183–196 (1987)
17. Shafer, G.: A Mathematical Theory of Evidence, Vol. 42. Princeton University Press (1976)
18. Smets, P.: Information content of an evidence. Int. J. Man-Mach. Stud. **19**(1), 33–43 (1983)
19. Yager, R.R.: Entropy and specificity in a mathematical theory of evidence. In: Classic Works of the Dempster-Shafer Theory of Belief Functions, pp. 291–310. Springer (2008)

Estimation of Conditional Value-at-Risk in Linear Model

Jana Jurečková[1(✉)], Jan Picek[2], and Jan Kalina[3,4]

[1] The Czech Academy of Sciences, Institute of Information Theory and Automation, CZ -182 00 Prague 8, Prague, Czech Republic
jureckova@utia.cas.cz

[2] Technical University in Liberec, Faculty of Science, Humanities, and Education, Liberec, Czech Republic
jan.picek@tul.cz

[3] Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
kalina@cs.cas.cz

[4] The Czech Academy of Sciences, Institute of Computer Science, Prague, Czech Republic

Abstract. The conditional value-at-risk (CVaR) represents a popular risk measure often exploited e.g. within portfolio optimization. The situation with a nuisance linear regression is considered here; in other words, we do not observe directly the loss Z of interest, but only $Y = \beta_0 + \mathbf{X}\boldsymbol{\beta} + Z$, where the covariates are not under our control. We propose a novel estimator of CVaR(Z) based on the *averaged two-step regression quantile* combined with an R-estimate of regression parameters.

Keywords: Linear regression · Risk measure · Conditional value-at-risk · Nuisance regression · Averaged regression quantile · Two-step regression quantile

1 Introduction

We follow the variables $\mathbf{Z}_n = (Z_1, \ldots, Z_n)^\top$ measuring the loss Z of an asset or of a portfolio at times 1, 2, . . . , n. They are assumed to be independent and identically distributed (*i.i.d.*) with a distribution function F, satisfying

$$\int z dF(z) = 0, \quad \int z^2 dF(z) < \infty, \tag{1}$$

otherwise unknown. We are interested in the possible risk of the loss Z in a given period and with a particular confidence level $\alpha \in (0, 1)$. The popular risk measure is the *Conditional Value-at-Risk* (CVaR), equal to

$$\mathsf{CVaR}_\alpha(\mathsf{Z}) = \mathbf{E}\{\mathsf{Z}|\mathsf{Z} > \mathsf{F}^{-1}(\alpha)\} = (1-\alpha)^{-1} \int_\alpha^1 \mathsf{F}^{-1}(\mathsf{t})\mathsf{dt} = (1-\alpha)^{-1} \int_{\mathsf{F}^{-1}(\alpha)}^\infty \mathsf{zdF(z)}. \tag{2}$$

Devoted to the Memory of Pranab Kumar Sen.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 200–207, 2024.
https://doi.org/10.1007/978-3-031-65993-5_24

It has obtained applications in many areas immediately after its introduction; let us mention the management of water supplies, risk management of the social security fund, the cash flow risk measurement for non-life insurance industry, the financial risk in the industrial areas, operational risk in the banks, and others. In the finance is CVaR taken as a popular a risk measure in portfolio optimization or management of extreme risk (cf. [5]). There is a rich bibliography on the subject, both from theoretical and applications points. Innovative ideas on the subject and estimation methods were presented e.g. in [1,20,21].

A review of estimation methods for CVaR, based on data, is collected in [18]. It describes the parametric, nonparametric and semiparametric estimation methods, supplemented with the computer software, partially available from the R package (*R* Development Core Team, 2012). In the parametric setup, [18] admits the skew and asymmetric probability distributions. The computing methods based on characteristic functions and Fourier analysis through the empirical data are studied in [22]. There is also compared the Value-at-Risk (VaR) and CVaR with the new *expectile based risk measure* (ERM); the ERM is found as the only risk measure that is both coherent and elicitable. Recently, [12] exploited the concept of pseudo-capacities, elaborated in [3,9], in estimating the CVaR in nonparametric and measurement error situations.

The risk of a loss is often affected by a nuisance regression caused by exogenous market variables as interest rates, overall market sentiment, liquidity shocks, etc. Effects of nuisance regression in testing predictive models were recently considered either in [17] for assessing value-at-risk or in [4] for combining two forecasts for assessing CVaR. Joint fitting VaR and CVaR in regression models by a two-step procedure has been recently considered in [8].

In the present paper, we consider estimating CVaR in the situation that the loss of interest is unobservable, being affected by the covariates within a nuisance linear regression model. Moreover, the situation is nonparameric, with an unknown probability distribution of the loss. The literature is almost void in estimation of CVaR for models with a nuisance regression. Mathematical solution of estimating CVaR under a nuisance linear regression was considered by Trindade et al. [21]; they proposed to estimate the nuisance regression parameters with an M-estimate, minimizing a convex criterion under the restriction that the resulting CVaR does not exceed a given $\eta > 0$. However, the M-estimator and the resulting CVaR are not scale equivariant, while the scale equivariance is a desirable property of measures in applications. Comparing with the proposal in [21], our novel estimator of CVaR proposed here is equal to the *averaged regression* α*-quantile* of residuals with respect to the R-estimate of the regression parameters. As such, the estimate of CVaR is scale-equivariant and its asymptotics follows from the asymptotics for R-estimators and averaged regression quantile.

Section 2 of this paper overviews regression quantiles together with averaged two-step regression quantiles, which are exploited in Sect. 3 to define a novel CVaR estimate for a nuisance regression.

2 Regression Quantile and Its Two-Step Version

2.1 Motivation and Notation

The averaged two-step regression α-quantile, introduced in [11], approximates the quantile function $F^{-1}(\alpha)$ of the errors asymptotically in probability for $n \to \infty$, up to the standardization with the regression parameters (cf. [14]). Because its number of breakpoints equals exactly to n, while in the case of ordinary regression quantile it is much larger, and it is nondecreasing in $\alpha \in (0,1)$, the averaged two-step regression quantile also facilitates joint fitting the VaR and CVaR in the regression model.

Even if distribution function of Z is unknown, the estimate of $\mathsf{CVaR}_\alpha(\mathsf{Z})$ in (2) can be obtained from the empirical quantile function of independent observations $Z_1, Z_2, \ldots, Z_n$. The resulting estimate would be

$$\widehat{\mathsf{CVaR}}_\alpha(Z) = \lfloor n(1-\alpha) \rfloor^{-1} \sum_{i=\lfloor n\alpha \rfloor}^{n} Z_{n:i} = \lfloor n(1-\alpha) \rfloor^{-1} \sum_{i=1}^{n} Z_i \, \mathcal{I}[Z_{i\geq} Z_{n:[n\alpha]}]. \quad (3)$$

However, the financial returns are often regressed on some covariates, and we can observe only the variables $\mathbf{Y}_n = (Y_{n1}, ..., Y_{nn})^\top$, what are the $\mathbf{Z}_n$ affected by covariates $\mathbf{X}_n$ with unknown intensities, measured by regression coefficients $\beta_1, \ldots, \beta_n$. Taking this into account, we work with the linear regression model

$$\mathbf{Y}_n = \beta_0 \mathbf{1}_n + \mathbf{X}_n \boldsymbol{\beta} + \mathbf{Z}_n \quad (4)$$

with observations $\mathbf{Y}_n = (Y_{n1}, \ldots, Y_{nn})^\top$, unknown parameters β_0 (intercept), $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ (scales), $\mathbf{1}_n = (1, \ldots, 1)^\top \in \mathrm{R}_n$ and the $n \times p$ matrix $\mathbf{X} = \mathbf{X}_n$ of covariates rows $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$.

Because $Z_1, \ldots, Z_n$ are not available, every inference on the loss Z and on F^{-1}, hence also estimating $\mathsf{CVaR}_\alpha(Z)$, are possible only by means of observations of Y. Thus we have to look for an alternative explicit estimating of $\mathsf{CVaR}_\alpha(\mathsf{Z})$. Pioneering ideas in estimating the CVaR were presented in [1,20,21]. Compared to them, we take as a main tool for estimating CVaR the two-step α-regression quantile of the model (4), originated and illustrated by the present authors in [11,13,14].

2.2 Regression Quantile

The model (4) can be rewritten as the model

$$Y_{ni} = \beta_0 + \mathbf{x}_{ni}^\top \boldsymbol{\beta} + Z_{ni}, \; i = 1, \ldots, n \quad (5)$$

with covariates $\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nn}$, each element of $\mathbb{R}_p$. For the sake of brevity, we also use the notation $\mathbf{x}_{ni}^* = (1, x_{i1}, \ldots, x_{ip})^\top$, $i = 1, \ldots, n$. Let $\widehat{\boldsymbol{\beta}}_n(\alpha) \in R_{p+1}$ be the α-regression quantile of model (4), $0 < \alpha < 1$, i.e. the solution of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - b_0 - \mathbf{x}_i^\top \mathbf{b}) = \min, \; b_0 \in \mathbb{R}_1, \; \mathbf{b} \in \mathbb{R}_p. \quad (6)$$

If derivative f of F exists and is positive in a neighborhood of the quantile $F^{-1}(\alpha)$, and if the matrix

$$\mathbf{Q}_n = n^{-1} \left(\mathbf{1}_n, \mathbf{X}_n\right)^\top \left(\mathbf{1}_n, \mathbf{X}_n\right)$$

is positively definite starting with some n, then $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_n(\alpha) - \check{\boldsymbol{\beta}}(\alpha))$ admits the asymptotic representation (see e.g. [15])

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_n(\alpha) - \check{\boldsymbol{\beta}}(\alpha)) = n^{-\frac{1}{2}}(f(F^{-1}(\alpha))^{-1}\mathbf{Q}_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i^*(\alpha - I[Z_i < F^{-1}(\alpha)]) + o_p(1) \tag{7}$$

as $n \to \infty$, where $\check{\boldsymbol{\beta}}(\alpha) = (F^{-1}(\alpha) + \beta_0, \beta_1, \ldots, \beta_p)^\top$ is the population counterpart of the regression quantile. The intercept part of the representation (7) is rewritten as

$$\begin{aligned} \hat{\beta}_{n0}(\alpha) - \beta_0 - F^{-1}(\alpha) &= (nf(F^{-1}(\alpha)))^{-1} \sum_{i=1}^{n} (\alpha - I[Z_i < F^{-1}(\alpha)]) + o_p(n^{-\frac{1}{2}}) \\ &= Z_{n:[n\alpha]} - F^{-1}(\alpha) + o_p(n^{-\frac{1}{2}}) \end{aligned} \tag{8}$$

as $n \to \infty$, where the first equality follows from (7), while the second equality follows from the Bahadur representation of sample quantile. $\hat{\beta}_{n1}(\alpha), \ldots, \hat{\beta}_{np}(\alpha)$ are consistent estimates of the slope parameters $\beta_1 \ldots, \beta_p$. The slope components of regression quantile are asymptotically independent of the intercept component $\hat{\beta}_{n0}(\alpha)$. The solution of (6) minimizes the $(\alpha, 1-\alpha)$ convex combination of residuals $(Y_i - \mathbf{x}_i^{*\top}\mathbf{b})$ over $\mathbf{b} \in \mathbf{R}^{p+1}$, where the choice of α depends on the balance between underestimating and overestimating the respective losses Y_i. The increasing $\alpha \nearrow 1$ reflects a greater concern about underestimating losses Y, comparing to overestimating. A useful functional of the regression quantile is the *averaged regression α-quantile*, the weighted mean of components of $\widehat{\boldsymbol{\beta}}_n(\alpha)$, $0 \leq \alpha \leq 1$:

$$\bar{B}_n(\alpha) = \overline{\mathbf{x}}_n^{*\top}\widehat{\boldsymbol{\beta}}_n(\alpha) = \widehat{\beta}_{n0}(\alpha) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}\widehat{\beta}_j(\alpha), \quad \overline{\mathbf{x}}_n^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^* \tag{9}$$

As shown in [11], the $\bar{B}_n(\alpha) - \beta_0 - \bar{\mathbf{x}}_n^\top\boldsymbol{\beta}$ is asymptotically equivalent to the $[n\alpha]$-quantile $e_{n:[n\alpha]}$ of the model errors, if they are identically distributed.

2.3 Two-Step Regression Quantile

The two-step regression quantile was introduced in [13] and later studied e.g. in [14], where it was shown that it is asymptotically equivalent to the ordinary α-regression quantile. The two-step regression α-quantile combines the rank-estimator (R-estimator) $\widetilde{\boldsymbol{\beta}}_{nR}$ of the slope components $\boldsymbol{\beta}$ with the $[n\alpha]$ order statistics of the residuals $Y_i - \mathbf{x}_i^\top\widetilde{\boldsymbol{\beta}}_{nR}$, $i = 1, \ldots, n$. The two-step regression quantile first estimates the slope components $\boldsymbol{\beta}$ by means of an R-estimate

$\widetilde{\boldsymbol{\beta}}_{nR}(\lambda) \in \mathbb{R}^p$, defined as a minimizer of the Jaeckel's measure of the rank dispersion [10] with a fixed $\lambda \in (0,1)$:

$$\sum_{i=1}^{n}(Y_i - \mathbf{x}_i^\top \mathbf{b})[a_i(\lambda, \mathbf{b}) - \bar{a}_n(\lambda)] = \min \tag{10}$$

with respect to $\mathbf{b} = (b_1, b_2, \ldots, b_p)^\top \in \mathbb{R}^p$. The notation in (10) means:

$$a_i(\lambda, \mathbf{b}) = \begin{cases} 0 & \ldots \quad R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b}) < n\lambda \\ R_i - n\lambda & \ldots \quad n\lambda \le R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b}) < n\lambda + 1 \\ 1 & \ldots \quad n\lambda + 1 \le R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b}), \end{cases}$$

Here $R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b}), i = 1, \ldots, n$ are the ranks of the residuals, and $a_i(\lambda, \mathbf{b})$ are known as *Hájek's rank scores* (see [7]). Note that $\bar{a}_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n} a_i(\lambda, \mathbf{b})$ is constant in $\mathbf{b}$, as an average of the rank scores. The minimization (10) can be rewritten as

$$\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \mathbf{b}\right) a_i(\lambda, \mathbf{b}) = \min . \tag{11}$$

It implies that the solution of (10) is invariant to the intercept, which is a nuisance component. The solution of (10) and (11) is the R-estimator $\widetilde{\boldsymbol{\beta}}_{nR}(\lambda)$ of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, generated by the following score function $\varphi_\lambda : (0,1) \mapsto \mathbb{R}^1$:

$$\varphi_\lambda(u) + (1-\lambda) = \begin{cases} 0 \ldots 0 \le u < \lambda \\ 1 \ldots \lambda \le u \le 1 \end{cases} \tag{12}$$

Generally, as the score function we can use another nondecreasing square integrable function on $(0,1)$. By [15], $\widetilde{\boldsymbol{\beta}}_{nR}(\lambda)$ consistently estimates $\boldsymbol{\beta}$ under the following conditions on F and on $\mathbf{X}_n$:

(F1) The distribution function F has a continuous density f with a positive an finite Fisher information $\mathcal{I}(f)$.

(X1) Assume that, as $n \to \infty$,

$$n^{-1}\mathbf{V}_n = O_p(1) \quad \text{where} \quad \mathbf{V}_n = \sum_{i=1}^{n}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)^\top, \tag{13}$$

$$\max_{1 \le i \le n} \|\mathbf{x}_{ni} - \bar{\mathbf{x}}_n\| = o(n^{1/4}), \quad \bar{\mathbf{x}}_n = n^{-1}\sum_{i=1}^{n}\mathbf{x}_{ni}.$$

Moreover, we assume that $\mathbf{V}_n$ satisfies

$$\lim_{n\to\infty} \max_{1 \le i \le n} (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)^\top \mathbf{V}_n^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) = 0. \tag{14}$$

Under conditions (F1) and (X1), the R-estimator $\widetilde{\boldsymbol{\beta}}_{nR} = \widetilde{\boldsymbol{\beta}}_{nR}(\lambda)$ admits the following asymptotic representation, as $n \to \infty$ (see e.g. [15] for the proof):

$$\widetilde{\boldsymbol{\beta}}_{nR} - \boldsymbol{\beta} = (f(F^{-1}(\lambda)))^{-1}\mathbf{V}_n^{-1}\sum_{i=1}^{n}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\Big(I[Z_{ni} > F^{-1}(\lambda)] - (1-\lambda)\Big) + o_p(n^{-1/2}), \tag{15}$$

hence $\|n^{1/2}(\widetilde{\boldsymbol{\beta}}_{nR}-\boldsymbol{\beta})\| = O_p(1)$. The intercept component of the two-step regression α-quantile is defined as the $[n\alpha]$-quantile of the residuals $Y_i - \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda)$, $i = 1, \dots, n$. Denote it as $\widetilde{\beta}_{nR,0}(\alpha)$, hence

$$\widetilde{\beta}_{nR,0}(\alpha) = \left(Y_i - \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda)\right)_{n:[n\alpha]}$$

and we define the two-step α-regression quantile as the vector in $\mathbb{R}_{p+1}$

$$\widetilde{\boldsymbol{\beta}}_n(\alpha) = \left(\widetilde{\beta}_{nR,0}(\alpha)(\widetilde{\boldsymbol{\beta}}_{nR}(\lambda))^\top\right)^\top. \tag{16}$$

Hence, the averaged two-step regression α-quantile equals to

$$\widetilde{B}_{n\alpha} = \widetilde{\beta}_{nR,0}(\alpha) + \bar{\mathbf{x}}_n^\top v_{nR}(\lambda) = \left(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda)\right)_{n:[n\alpha]} \tag{17}$$

It has been introduced in [14], where it is proven that

$$\widetilde{B}_{n\alpha} - \beta_0 - \bar{\mathbf{x}}_n^\top \boldsymbol{\beta} = Z_{n:[n\alpha]} + o_p(n^{-1/2}) = F^{-1}(\alpha) + o_p(n^{-1/2}) \ \text{ as } \ n \to \infty \tag{18}$$

uniformly for $\alpha \in (\varepsilon, 1-\varepsilon)$, $0 < \varepsilon \le 1/2$, and for any fixed $\lambda \in (0,1)$.

3 Estimation of CVaR_α

If there are available independent observations $Z_1, Z_2, \dots, Z_n$, then $\mathsf{CVaR}_\alpha(\mathsf{Z})$ in (2) can be estimated with the aid of their empirical quantile function, even if distribution function of Z is unknown. The corresponding estimate of the conditional value-at-risk of Z would be

$$\widehat{\mathsf{CVaR}}_{n\alpha}(Z) = \lfloor n(1-\alpha)\rfloor^{-1} \sum_{i=\lfloor n\alpha\rfloor}^{n} Z_{n:i} = \lfloor n(1-\alpha)\rfloor^{-1} \sum_{i=1}^{nZ_i} I[Z_i \ge Z_{n:[n\alpha]}]. \tag{19}$$

If the observations $Z_1, Z_2, \dots, Z_n$, are not at disposal, we can profit from approximations (17) and (18) and estimate the CVaR_α by means of the averaged two-step regression quantile $\widetilde{B}_n(\alpha)$, which explicitly contains only observations $Y_1, \dots, Y_n$. The resulting estimator is

$$\widehat{\mathsf{CVaR}}_{n\alpha}(Y) = \lfloor n(1-\alpha)\rfloor^{-1} \sum_{\alpha \le \delta < 1} \widetilde{B}_n(\delta) - \beta_0 - \bar{\mathbf{x}}_n^\top \boldsymbol{\beta}. \tag{20}$$

Hence, the estimate is determined up to the standardization with nuisance $\beta_0 + \bar{\mathbf{x}}_n^\top \boldsymbol{\beta}$, what can be approximated by means of $Y_{n1}, \dots, Y_{nn}$ under the conditions (F1) and (X1), namely by $\bar{\mathbf{x}}_n^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda)$ and by $\bar{Y}_n$. Indeed, by (5) and (1),

$$\bar{Y}_n = \beta_0 + \bar{\mathbf{x}}_n^\top \boldsymbol{\beta} + \bar{Z}_n = \beta_0 + \bar{\mathbf{x}}_n^\top \boldsymbol{\beta} + o_p(n^{-1/2})$$

as $n \to \infty$.

Summarizing, our proposed estimate of the $\mathsf{CVaR}_\alpha(\mathsf{Z})$ in the linear model is described in the following theorem:

Theorem 1. *Assume that the distribution of Z satisfies condition (F1), let $\int z dF(z) = 0,\ 0 < \int z^2 dF(z) < \infty$, and let the matrix $\mathbf{X}_n$ satisfy (X1). Then*

$$\widehat{CVaR}_{n\alpha}(Y) = \lfloor n(1-\alpha) \rfloor^{-1} \sum_{\delta:\alpha \le \delta < 1} \Big(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda) \Big)_{n:[n\delta]} - \bar{Y}_n \quad (21)$$

is a $\sqrt{n}$-consistent estimate of $\mathsf{CVaR}_\alpha(\mathsf{Z})$ for any fixed $\lambda \in (0,1)$; more precisely

$$\widehat{CVaR}_{n\alpha}(Y) = \mathsf{CVaR}_\alpha(Z) + o_p(n^{-1/2}) \ \ as \ \ n \to \infty. \quad (22)$$

for any $\alpha \in (0,1)$. Moreover,

$$\Big(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}(\lambda) \Big)_{n:[n\alpha]} - \bar{Y}_n = \mathsf{VaR}_\alpha(\mathsf{Z}) + \mathsf{o_p}(\mathsf{n}^{-1/2}) \ \ as \ \ \mathsf{n} \to \infty. \quad (23)$$

Conclusion

A novel nonparametric estimate of the Conditional Value-at-Risk of variable Z is proposed here in the situation that the observations are the responses under the presence of nuisance regression. The estimate is based on the averaged two-step regression α-quantile of the linear model, exploiting an R-estimator of the slope components. As such, the estimates are $\sqrt{n}$ consistent and scale equivariant estimates of the Conditional Value-at-Risk of variable Z, in spite of the presence of the nuisance regression.

The novel estimators may find applications not only in finance, but also in various other (non-financial) models, even in situations under the presence of nuisance covariates. This can be of interest e.g. in the risk assessment in metrology [2,16] or within integrated energy systems for optimal scheduling the energy supply for large cities [19]). The novel tools come to hand, if the risk is influenced by a combination of more nuisance effects, including weather conditions (temperature, humidity, wind speed, precipitation, seasonal trends), energy price fluctuations, natural events (power outages), or industrial activity.

Acknowledgements. The authors thank the Editor and Referees for their attentive reading the text and for their comments. The research was supported by the grants 22-03636S (J. Jurečková, J. Picek) and 24-11146S (J. Kalina) of the Czech Science Foundation.

References

1. Bassett, G.W., Koenker, R., Kordas, W.: Pessimistic portfolio allocation and Choquet expected utility. J. Financ. Econ. **2**, 477–492 (2004)
2. Božić, D., Runje, B., Razumić, A.: Risk assessment for linear regression models in metrology. Appl. Sci. **14**, 260 (2024)
3. Buja, A.: On the Huber-Strassen theorem. Probab. Theory Rel. **73**, 149–152; Ann. Oper. Res. **313**, 755–772 (1986)

4. Dimitriadis, T., Liu, X., Schnaitmann, J.: Encompassing tests for value at risk and expected shortfall multistep forecasts based on inference on the boundary. J. Financ. Econ. **21**, 412–444 (2023)
5. Dixit, V., Tiwari, M.K.: Project portfolio selection and scheduling optimization based on risk measure: a conditional value at risk approach. Ann. Oper. Res. **285**, 9–33 (2020)
6. Gneiting, T.: Making and evaluating point forecasts. J. Am. Stat. Assoc. **106**, 746–62 (2011)
7. Hájek, J.: Extension of the Kolmogorov-Smirnov test to regression alternatives. In: LeCam, L. (ed.) Proceedings of Bernoulli-Bayes-Laplace Seminar, pp. 45–60. University of California Press (1965)
8. He, X., Tan, K.M. Zho, W.-X.: Robust estimation and inference for expected shortfall regression with many regressors. J. R. Stat. Soc. Ser. B Stat. Method. **85**(4), 1223–1246 (2023)
9. Huber, P., Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Statist. **2**, 251–273 (1973)
10. Jaeckel, L.A.: Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist. **43**, 1449–1459 (1972)
11. Jurečková, J., Picek, J.: Averaged regression quantiles. In: Lahiri, S. et al., (eds) Contemporary Developments in Statistical Theory pp. 203–216. (Chapter 12). Springer, Switzerland (2014)
12. Jurečková, J., Kalina, J., Večeř, J.: Estimation of expected shortfall under various experimental conditions (2022). arXiv:2212.12419v1
13. Jurečková, J., Picek, J.: Two-step regression quantiles. Sankhya A **67**Part 2, 227–252 (2005)
14. Jurečková, J., Picek, J., Schindler, M.: Empirical regression quantile processes. Appl. Math. **65**, 257–269 (2020)
15. Jurečková, J., Sen, P.K., Picek, J.: Methodology in Robust and Nonparametric Statistics. CRC Press, Boca Raton, FL (2013)
16. Kalina, J., Vidnerová, P., Soukup, L.: Modern approaches to statistical estimation of measurements in the location model and regression. In: Aswal, D.K., et al. (eds.) Handbook of Metrology and Applications, pp. 2355–2376. Springer, Singapore (2023)
17. Khalaf, L., Leccadito, A., Urga, G.: Multilevel and tail risk management. J. Financ. Econ. **20**, 839–874 (2022)
18. Nadarajah, S., Zhang, B., Chan, S.: Estimation methods for expected shortfall. Quant. Finance **14**(2), 271–291 (2013)
19. Padhan, H., Padhang, P.C., Tiwari, A.K., Ahmed, R., Hammoudeh, S.: Renewable energy consumption and robust globalization(s) in OECD countries: do oil, carbon emissions and economic activity matter? Energy Strateg. Rev. **32**, 100535 (2020)
20. Pflug, G.C.: Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev, S.P. (ed.) Probabilistic Constrained Optimization Methodology and Applications. Springer, New York (2000)
21. Trindade, A.A., Uryasev, S.P., Shapiro, A., Zrazhevsky, G.: Financial prediction with constrained tail risk. J. Bank. Finance **31**, 3524–3538 (2007)
22. Zaevski, T.S., Nedeltchev, D.C.: From BASEL III to BASEL IV and beyond: expected shortfall and expectile risk measures. Int. Rev. Fin. Anal. **87**, 102645 (2023)

Common Models of Errors in Variables

Henrik Kaiser(✉)

Paris Lodron University of Salzburg, Salzburg, Austria
henrik.kaiser@plus.ac.at

Abstract. Blurred observations can be described in multiple ways. If the errors occur as a scaling or shifting variable, we find ourselves in the frequently employed multiplicative or additive model of errors in variables. In these, the determination of the distribution or density function of the unblurred variable, under the assumption of independent errors, corresponds to some kind of deconvolution problem. We briefly present both models and point out their parallels. Special focus lies on the additive model, which was first studied in the late 1980 s and is still subject of further research. After a discussion of the traditional approach by virtue of Fourier transforms, we outline its developments until today. A distinction between known and unknown errors is made.

Keywords: Errors in variables · Measurement errors · Additive · Multiplicative · Deconvolution · Unobservable data · Imprecise data · Blurred data · Noise · ill-posed · Inverse problems · Integral transforms

1 Shades of Measurement Errors

Since Kolmogorov[1] it is well-known that the distribution function (d.f.) F of some random variable X consistently can be estimated through the empirical distribution function (e.d.f.)

$$F_n(\xi) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{X_k \leq \xi\}, \qquad \xi \in \mathbb{R}, \tag{1}$$

provided a sample of independent, identically distributed (i.i.d.) observations $X_1, \ldots, X_n \sim F$ of size $n \in \mathbb{N}$ is available. Unfortunately, this is often not the case and rather than to X itself, we merely have access to a blurred version Y. There are several possibilities to describe the dependence of Y on X. For instance, if ε denotes another random variable, associated with an error or a noise, by supposing that

$$Y = q(X, \varepsilon), \tag{2}$$

[1] See [16] for the original Italian paper and [23] for a summary in English.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 208–216, 2024.
https://doi.org/10.1007/978-3-031-65993-5_25

for a specific function $q : \mathbb{R}^2 \to \mathbb{R}$. Of course, in practice, it is only clear that such a function q exists but not what it looks like. Also, generally, X and ε may depend on each other. In the sequence, we assume $\varepsilon \sim H$ and $Y \sim G$, for arbitrary d.f. H and G. Moreover, if existing, we denote their probability density functions (p.d.f.) by the corresponding small letters f, g and h. Now, the most intuitive examples for functions q act by componentwise multiplication and addition, viz

$$Y = X\varepsilon, \tag{3a}$$

$$Y = X + \varepsilon, \tag{3b}$$

respectively referred to as the multiplicative and the additive model of errors in variables. Actually, by exponentiation, (3b) turns out as a special case of (3a). Conversely, (3a) admits a logarithmic transformation to (3b) only if $X, \varepsilon > 0$ almost surely. Yet, a separate discussion of both models can be helpful, as the characterization of exponentially or logarithmically transformed random variables is possibly less convenient. In the multiplicative model, the observable variable Y can be conceived as a scaled or weighted version of the actual variable of interest X. In the additive model, the noise ε inflicts a shift. Basically, with regard to (3), given a Y-sample, there are two relevant problems. On the one hand, one may be interested in the associated unobservable realizations of X. On the other hand, one may want to estimate F (or f). In the last case, an appropriate starting point can be an integral representation of G. Concerning the multiplicative model of errors in variables, it requires a distinction between different signs of X and ε. Particularly if both variables are independent and almost surely positive,

$$G(\xi) = \int_0^\infty F(\xi z^{-1}) H(dz), \qquad \xi \in \mathbb{R}, \tag{4}$$

corresponding to an integral of *Mellin (or multiplicative) convolution* type. Similarly, in the additive model of errors in variables, for arbitrary independent X and ε, the integral representation of G is the *(additive) convolution* of F and H, i.e.,

$$G(\xi) = \int_{-\infty}^\infty F(\xi - z) H(dz), \qquad \xi \in \mathbb{R}. \tag{5}$$

A more common notion for these integrals in older literature is (Stieltjes) *resultant* or the German word *Faltung*, see p. 51–52 in [25], p. 84 in [28] and p. 71 in [29]. Usually, if the context is clear, one simply speaks of convolution and writes $G = F * H$. Accordingly, in both models the recovery of F is called *deconvolution*. From an operator theoretical point of view, deconvolution is considered an *ill-posed* inverse problem, due to [13], essentially because of the unboundedness of the respective inverse operator. Notice that this is typical for integral

equations of the first kind, of which the above equations indeed are. However, in operator theory, convolution acts on function spaces, i.e., it is a *global* operation, whereas the deconvolution problem is rather *local*, since it amounts to solve a specific equation. From a stochastic perspective, deconvolution increases in difficulty because some quantities can only be estimated, especially G (and sometimes also H, depending on the assumptions). But even then, it is generally *not* ill-posed if we are interested in d.f., instead of p.d.f.. As a matter of fact, density estimation *is* ill-posed already in an error-free setup.

While deconvolution is a rather modern terminology, the actual problem has a long tradition in calculus, where it was examined long before its stochastic treatment, cf. Ch. XI in [25], §1.9 in [26] or Ch. V, §8 in [28]. See also §4.5 in [19] for an overview. Integrals of convolution-type frequently occur in the theory of integral transforms, simply because a suitable transformation leads to a multiplicative decomposition. To be exact, (4) and (5) become a product in the Mellin and Fourier domain,[2], respectively. This is the reason, why also many statisticians pursue deconvolution from this direction. For example, Mellin transforms were employed in [3,4] for the non-parametric estimation of the target p.d.f. f in the presence of multiplicative errors. In [5] an additional treatment of the survival function is included. Moreover, by similar means, the estimation of F was conducted in [18]. Of course, there also exist different approches, like a maximum likelihood technique by [27]. However, whereas deconvolution with additive errors enjoys wide popularity, the literature on multiplicative errors is apparently much sparser. A comparison exposes many parallels between the techniques that rely on Mellin transforms and methods known from Fourier estimators. A possible explanation is, despite the functions defined by these integrals in theory exhibit a fairly different structure, that one faces similar issues when it comes to estimation, namely the loss of smoothness properties. A common way to reobtain these is the use of smoothing kernels. We will explain this below in detail for the additive model of errors in variables, with independent errors, including a list of recent results. After its first consideration in the late 1980 s, the deconvolution problem regained much attention in the first decade of the 21st century through many contributions. The interest in this research area during the last years seems to be on a constant ordinary level with a lack of novel ideas. Yet, the problem remains unsolved as there still is no satisfying general approach. At the same time, it is highly relevant in our world, where data plays an increasing role. The variable Y can be conceived in various ways, for instance as an imprecise or blurred measurement, due to certain circumstances, or even as intentionally faked data.

[2] Recall that $\mathbb{E}\left[X^{t-1}\right] = \int_0^\infty x^{t-1} F(dx)$ is the Mellin-Stieltjes transform of F. A necessary condition for its existence, for some $t \in \mathbb{R}$, is that $X > 0$ almost surely. Furthermore, $\Phi_X(t) := \mathbb{E}\left[e^{itX}\right] = \int_{-\infty}^\infty e^{itx} F(dx)$ constitutes the Fourier-Stieltjes transform of F also known as the characteristic function of X. It establishes a uniformly bounded function of $t \in \mathbb{R}$ for any real-valued random variable X. Both transforms are invertible.

For a more comprehensive overview on measurement errors in statistics, we recommend the monographs [17,30], which also cover some shades of errors in the context of regression and beyond.

2 The Additive Model of Errors in Variables

Consider the additive model of errors in variables with independent X and ε. As we mentioned above, convolution then corresponds to multiplication in the Fourier domain. To be exact, (5) implies that

$$\Phi_Y = \Phi_X \Phi_\varepsilon. \tag{6}$$

Note that actually (5) and (6) are equivalent. The last equation uniquely specifies Φ_X only if $\Phi_\varepsilon(t) \neq 0$ for almost every $t \in \mathbb{R}$, in which event one obtains that $\Phi_X = \{\Phi_\varepsilon\}^{-1}\Phi_Y$ almost everywhere and else by continuity. From this, with the aid of an inversion formula, one readily returns to F. The preceding arguments become invalid if there exists $(a, b) \subset \mathbb{R}$ with $\Phi_\varepsilon(t) = 0$ for any $a < t < b$. Then, $\Phi_X(t)$ is indeterminable, for all $a < t < b$, and hence F is eventually unidentifiable. That's the situation in calculus, where Φ_Y and Φ_ε are known. However, in stochastics Φ_Y is only estimable. A consistent estimator is the *empirical characteristic function*

$$\Phi_Y(t, n) := \frac{1}{n} \sum_{k=1}^{n} e^{itY_k}, \qquad t \in \mathbb{R}, \tag{7}$$

for an i.i.d. sample $Y_1, \ldots, Y_n \sim G$. It is almost periodic with random zeros and especially non-decaying as $t \to \pm\infty$. Moreover, [10] ascertained almost sure uniform convergence of $\Phi_Y(\cdot, n)$ to Φ_Y in any compact subset of $\mathbb{R}$, which fails to hold on the whole real axis, except for discrete F. The particular issues that may occur upon defining an estimator for Φ_X through (6), with Φ_Y replaced by $\Phi_Y(\cdot, n)$, depend on whether or not the error distribution is known. Roughly speaking, it must be expected that the resulting estimator is unbounded.

2.1 Known Error Distribution

The assumption of a known error distribution is equivalent to full information on the properties of the characteristic function Φ_ε. The target d.f. F is then uniquely identifiable, unless $\Phi_\varepsilon = 0$ on a non-zero set. This scenario was first treated in 1990 by [22], with regard to density estimation. Notice that density estimation in an error-free setup became popular already around the 1960 s through [20,21]. At that time, in case of observable X, it was first shown that an unbiased estimator for the p.d.f. f does *not* exist, thereby associating density estimation with the class of statistically ill-posed problems. This result is immediately connected with the piecewise continuity of the e.d.f., due to which its derivative is almost everywhere zero. But actually it is not a surprise, because how could we expect

general existence of an unbiased estimator for f, if we know that f need not exist at all. To fix both of these issues, use of a smoothing kernel was proposed. The approach by [22] to density deconvolution, i.e., to estimate the p.d.f. in the additive model of errors in variables, is actually quite similar. The authors adopted the technique from calculus and suggested to improve the vulnerability of the estimator for the characteristic function of Φ_X, viz of $\{\Phi_\varepsilon\}^{-1}\Phi_Y(\cdot, n)$, by means of an appropriate smoothing kernel. In this fashion, they received for f the estimator

$$\mathfrak{f}_n(\xi, \lambda_n) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi t} \frac{\Phi_Y(t, n)}{\Phi_\varepsilon(t)} \Phi_I(\lambda_n t) dt, \qquad \xi \in \mathbb{R}, \tag{8}$$

with bandwidth $\lambda_n > 0$ and a smoothing kernel, whose Fourier transform Φ_I assures absolute convergence of the integral, for a given fixed error distribution. Similar to the density estimator in an error-free environment, the choice of the kernel essentially affects the performance of $\mathfrak{f}_n(\cdot, \lambda_n)$. Besides, admissible Φ_I are as well determined by Φ_ε. Lastly, the bandwidth λ_n is supposed to tend to zero as $n \to \infty$, however, sufficiently slow in order to achieve a certain kind of convergence. A simple criterion for consistency was provided by [22] as well as a brief discussion of the mean integrated square error (MISE). They confined to non-vanishing Φ_ε and observed that the validity of their statements essentially depends on the tail behaviour of Φ_ε. Accordingly, the rate of convergence decreases for errors with rapidly decaying characteristic functions, thereby associating normally distributed errors with a worst case scenario. A more detailed study of the asymptotic behaviour of the density estimator was accomplished in [8,9], upon categorizing error distributions through the decay of $\Phi_\varepsilon(t)$ as $t \to \pm\infty$. Particularly in [9] the classes of *ordinary smooth* (algebraically decaying) and *super smooth* (exponentially decaying) functions were introduced, which still nowadays is used for reference, cf. [1,24,31]. However, this distinction is far from complete, as it only covers a few absolutely continuous distributions. Attempts to construct estimators with the capability to cover a larger scope of distributions, and especially to allow for arbitrary Φ_ε, apparently do not yet exist. Instead, methods that aim for a broader applicability often involve assumptions that merely hold for special distributions. See [2,11], for example. Notice that these results were presented in terms of the bilateral Laplace transform with imaginary argument, which is basically equivalent to the Fourier transform.

Finally, although Fourier methods are predominant, there exist alternative deconvolution techniques. Occasional use is made of the wavelet transform, see [6] or §2.2.2 in [17]. However, it relies on Hilbert spaces and thus requires $f \in L^2(\mathbb{R})$. Besides, the construction of a wavelet estimator is very elaborate and its performance is naturally determined by the respective orthonormal base. In [12] a maximum likelihood method with Bernstein polynomials was employed. An interesting approach that seems to lead towards a generalized direction can be found in [14] and §2.2.3 in [17], and it involves a ridge-parameter. In this, Φ_X is

estimated through

$$\Phi_\rho(t,n) := \frac{\Phi_\varepsilon(-t)\Phi_Y(t,n)}{\max\{\rho_n(t), |\Phi_\varepsilon(t)|^2\}},$$

with ridge-parameter or -function $\rho_n(t) > 0$. Note, if $\rho_n(t) < |\Phi_\varepsilon(t)|^2$, that the right hand side equals $\{\Phi_\varepsilon(t)\}^{-1}\Phi_Y(t,n)$, which is in fact unbiased with respect to $\Phi_X(t)$. In this regularization method, use of a kernel is optional. The remarkable property of $\Phi_\rho(\cdot,n)$ is the applicability to any kind of distribution. At the same time, one can even achieve absolute convergence of the associated estimator for F, represented as a Fourier-integral. Unfortunately, this approach received only little attention.

A possible reason, why there still is no widely applicable and promising approach to deconvolution, even after 30 years of research, is its predominant discussion with regard to density estimation, which already is a nasty problem without measurement errors. A reconsideration of the deconvolution problem on the level of d.f., in contrast, avoids some difficulties. For instance, if X and ε are discretely distributed with support $\mathbb{N}_0$, elementary manipulations show that

$$F(\xi) = \sum_{l=0}^{\lfloor \xi \rfloor} G(\xi - l) \sum_{j=0}^{l} (-1)^j \{p_\varepsilon(0)\}^{-j-1} p(j,l), \qquad \xi \in \mathbb{R}, \tag{9a}$$

where $p_\varepsilon(k) := \mathbb{P}(\varepsilon = k)$, for $k \in \mathbb{N}_0$, and

$$p(j,l) := \begin{cases} \sum\limits_{z_1,\ldots,z_j \in \mathbb{N}} \prod\limits_{i=1}^{j} p_\varepsilon(z_i), & \text{for } j \geq 1, \\ \delta_{\{0\}}(\{l\}), & \text{for } j = 0, \end{cases} \tag{9b}$$

with $\delta_{\{0\}}$ denoting the Dirac-measure at the origin. In addition, analogous representations for F can be found if X and ε are associated with more general unilateral distributions, one of them remaining discrete. Accordingly, deconvolution of d.f. need *not* be ill-posed. Further investigations, that are to be published, eventually facilitated the construction of a biased estimator for F that is determined solely through the available quantities $\Phi_Y(\cdot,n)$ and Φ_ε and does *not* involve any artificial smoothing kernels. It results from a sophisticated rearrangement of the product $\{\Phi_\varepsilon\}^{-1}\Phi_Y(\cdot,n)$, aiming at a cancellation whenever the reciprocal becomes unbounded. This novel estimator, called the *empirical deconvolution function*, possesses a representation as an absolutely convergent Fourier integral under weak assumptions.

2.2 Unknown Error Distribution

It is clearly debatable, how realistic in applications the assumption of a known error distribution is. Of course, there are scenarios, in which we have this information, but in most circumstances it is incomplete and H needs to be determined

separately. Nevertheless, a study of deconvolution with known errors in before is fundamental to properly assess the extent of the problem and its difficulties. Now, in case of an unknown H, additional caution is required with regard to the identifiability of F and H. To make this statement more precise, we assume the existence[3] of characteristic functions Φ_1 and Φ_2 with $\Phi_Y = \Phi_1\Phi_2$. It is then not immediately ascertainable, which factor corresponds to F and H, respectively. Moreover, there is no unique way to address this issue but it depends on the extent of missing information on H in each particular scenario and on the applicable means. For a selection see [1,15] and §2.6 in [17]. A rather mild case occurs if H is known to be associated with a certain parametric family of distributions, whose parameters remain to be estimated. In any case, if there is only partial or even no information on H, additional data is required. Ideally, a sample of the error variable could be generated independently from the Y-sample, but this seems unrealistic in practice. A more justifiable assumption is the availability of replicated measurements, i.e., $Y_{n,j_n} = X_n + \varepsilon_{n,j_n}$, with $X_n \sim F$ and $\varepsilon_{n,j_n} \sim H$, for $1 \le j_n \le J_n$ and $1 \le n \le N$. In this event, one can estimate the symmetrized error distribution $\bar{H}(\xi) := \int_{-\infty}^{\infty} H(\xi + z)H(dz)$ and its characteristic function $|\Phi_\varepsilon|^2$ through the differences $\Delta Y_n := Y_{n,j_{n_1}} - Y_{n,j_{n_2}}$, with $j_{n_1} \neq j_{n_2}$. In the special case $\Phi_\varepsilon > 0$, this indeed leads to an estimator for Φ_ε and the target d.f. F is identifiable, while H itself is only identifiable up to a shift. Lastly, an interesting and completely different approach to estimate F without additional data was presented by [7]. In this, the target d.f. F is characterized through the variance, motivated by what the authors call the random-universe principle. Their aim is the approximation of F by a discrete distribution, whose atoms can be received from minimizing a certain functional. However, the assumptions on the errors are fairly restrictive.

Acknowledgements. The author gratefully acknowledges the support of the WISS 2025 project 'IDA-lab Salzburg' (Land Salzburg, WISS 2025, 20102_F2300464-KPZ, 20204-WISS/225/348/3–2023).

References

1. Adusumilli, K., Kurisu, D., Otsu, T., Whang, Y.: Inference on distribution functions under measurement error. J. Econom. **215**(1), 131–164 (2020)
2. Belomestny, D.: Density deconvolution under general assumptions on the distribution of measurement errors. Ann. Statist. **49**(2), 615–649 (2021)
3. Belomestny, D., Goldenshluger, A.: Nonparametric density estimation from observations with multiplicative measurement errors. Ann. Inst. H. Poincare-Probab. Statist. **56**(1), 36–67 (2020)
4. Brenner Miguel, S., Comte, F., Johannes, J.: Spectral cut-off regularisation for density estimation under multiplicative measurement errors. Electronic J. Statist. **15**(1), 3551–3573 (2021)

[3] Notice that Φ_1 and Φ_2 generally are not unique. Indeed, also $\Phi_Y = \Phi_3\Phi_4$ with $\Phi_3(t) := e^{i\mu t}\Phi_1(t)$ and $\Phi_4(t) := e^{-i\mu t}\Phi_2(t)$ for every $\mu \in \mathbb{R}$. This example corresponds to a shift in the domain of random variables, i.e., $Y = (X + \mu) + (\varepsilon - \mu)$.

5. Brenner Miguel, S., Johannes, J., Siebel, M.: Multiplicative deconvolution under unknown error distribution. Freely accessible on arXiv (2023)
6. Cao, K., Zeng, X.: A data-driven wavelet estimator for deconvolution density estimations. Result. Math. **78**(4), 156 (2023)
7. Delaigle, A., Hall, P.: Methodology for non-parametric deconvolution when the error distribution is unknown. J. R. Statist. Soc., Ser. B, Statist. Methodol. **78**(1), 231–252 (2016)
8. Fan, J.: On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. **19**(3), 1257–1272 (1991)
9. Fan, J.: Asymptotic normality for deconvolution kernel density estimators. Sankhyā **53**(1), 97–110 (1991)
10. Feuerverger, A., Mureika, R.: The empirical characteristic function and its applications. Ann. Statist. **5**(1), 88–97 (1977)
11. Goldenshluger, A., Kim, T.: Density deconvolution with non-standard error distributions: rates of convergence and adaptive estimation. Electronic J. Statist. **15**(1), 3394–3427 (2021)
12. Guan, Z.: Fast nonparametric maximum likelihood density deconvolution using Bernstein polynomials. Stat. Sin. **31**(2), 891–908 (2021)
13. Hadamard, J.: Sur les dérivées des fonctions de lignes. Bull. Soc. Math. Fr. **30**, 40–43 (1902)
14. Hall, P., Meister, A.: A ridge-parameter approach to deconvolution. Ann. Statist. **35**(4), 1535–1558 (2007)
15. Kato, K., Sasaki, Y.: Uniform confidence bands in deconvolution with unknown error distribution. J. Econom. **207**(1), 129–161 (2018)
16. Kolmogorov, A.: Sulla determinazione empirica di una legge di distribuzione. G. Ist. Ital. Attuari **4**, 83–91 (1933)
17. Meister, A. (Ed.): Deconvolution Problems in Nonparametric Statistics. Lecture Notes in Statistics, vol. 193, Springer, Berlin, Heidelberg (2009)
18. Morozova, E., Belomestny, D., Panov, V.: Statistical inference for scale mixture models via Mellin transform approach. Statistics **58**(1), 209–229 (2024)
19. Paris, R., Kaminski, D.: Asymptotics and Mellin-Barnes Integrals. Encyclopedia of Mathematics and its Applications, vol. 85, Cambridge University Press (2001)
20. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. **33**(3), 1065–1076 (1962)
21. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. **27**(3), 832–837 (1956)
22. Stefanski, L., Carroll, R.: Deconvoluting kernel density estimators. Statistics **21**(2), 169–184 (1990)
23. Stephens, M.: Introduction to Kolmogorov (1933) on the empirical determination of a distribution. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics II: Methodology and Distribution, pp. 93–105. Springer Series in Statistics. Springer, New York, NY (1992)
24. Thuy, L., Phuong, C.: Deconvolution problem of cumulative distribution function with heteroscedastic errors. J. Korean Stat. Soc. **52**(2), 330–360 (2023)
25. Titchmarsh, E.C.: Introduction to the Theory of Fourier Integrals. Oxford Clarendon Press (1937)
26. Tricomi, F.: Integral Equations. Pure and Applied Mathematics, vol. 5, Dover Publications (1985 (1957))
27. Vardi, Y.: Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. Biometrika **76**(4), 751–761 (1989)

28. Widder, D.: The Laplace Transform. Princeton University Press (1946)
29. Wiener, N.: The Fourier Integral and Certain of its Applications. Cambridge University Press (1988 (1933))
30. Yi, G.Y., Delaigle, A., Gustafson, P.: Handbook of Measurement Error Models. Chapman and Hall/CRC (2021)
31. Zhu, W., Sun, S., Wang, P., Khakurel, H.: Asymptotic normality of the deconvolution kernel density estimator based on independent right censored data. Matematica **1**(2), 531–550 (2022)

Data Augmentation with ChatGPT for Assessing Subject Alignment

Louisa Kontoghiorghes(✉) and Ana Colubi

King's College London, London, UK
{louisa.1.kontoghiorghes,ana.colubi}@kcl.ac.uk

Abstract. As a statistical tool, topic modeling requires replication. Topics are identified as distributions over a set of words, and documents are described as mixtures of latent topics. Sometimes, text information is available through a unique document that covers well-defined subjects, but without replication, the intuitive representation of the document as a mixture of latent topics cannot be derived. Nevertheless, the available document can potentially generate more knowledge (not yet available) that could assist in employing compelling text mining tools. The proposal is to use ChatGPT to mimic the process of generating such knowledge. To illustrate the approach, a research proposal is used as the initial document. The aim is to verify if a given piece of research aligns with the subjects of the research proposal or not combining text mining and statistical tools. A second case-study, analyse Novel's chapters alignment with the Gothic genre.

Keywords: Text mining · Pre-trained transformer · Tf-idf · Hypothesis testing · Data augmentation

1 Introduction

Text analysis has made significant advancements in recent years with the integration of the Large Language Model (LLM). The exploitation of text documents to determine various behaviors, turning points, contexts, patterns, and sentiments has proven to have a significant impact [8–10].

Subject alignment has been used to assess domain relevance and similarities [4,5]. However, when assessing subject alignment using a reference document, the manual validation of topic labels can be a time-consuming process, leading to challenges in identifiability of topics [4,5]. Since topics are normally defined as latent variables, in order to uncover them a statistical text data sample is required. Sometimes, text data can be scarce due to rare cases of data, privacy protections, legal restrictions, and cost of data [1,6,12,15]. To address this limitation, the implementation of generative pre-trained transformer (GPT) models, e.g., deep neural network models, can assist in augmenting textual data based on a reference text [2,16,17]. These models introduce noise, which may include

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 217–224, 2024.
https://doi.org/10.1007/978-3-031-65993-5_26

content that is occasionally unrelated or unrealistic to the topic, but making them more robust [11,13].

In this paper, a methodology for subject alignment with scarce data is considered. The proposal is to generate text data with ChatGPT. In detail, the subjects are here determined by a set of keywords solving the identification problem. To mimic real-life text, additional related noise is added to the model. That is, ChatGPT is employed to generate a sample of documents that preserve a part of the subjects discussed in a reference document d_0 and combine them with related subjects, not necessarily in d_0, but which could be naturally found together in real life. To achieve this, the document keywords are combined with terms from an external related list of keywords to generate noise and realistic statistical variability. The sample of simulated documents is used to create a reference of what is considered aligned with document d_0. The relative total term-frequency-inverse-document-frequency (tf-idf) of a document is suggested as an indicator of its alignment with d_0 [14]. Statistical outlier detection methods can be applied to assess the alignment of a given document. In this paper, a box plot serves as a visual aid, while the Z-score is employed to test the alignment of the given document with d_0.

The methodology is applied in a scientific context on three documents, including the abstract of this paper, to assess the subject alignment with the topics of discussion of the COST Action HiTEc (http://www.hitecaction.org/, Text, functional and other high-dimensional data in econometrics: New models, methods, applications). A second case study is conducted to analyse the Chapters of Jane Austen's Novel Northanger Abbey, in alignment with the Gothic genre.

The rest of the paper is as follows. A brief overview of topic modeling is provided in Sect. 2. In Sect. 3, the proposed algorithm using data augmentation with ChatGPT is detailed. In Sect. 4, the methodology is illustrated through the HiTEc application and Novel's Chapters. Finally, the conclusion is discussed in Sect. 5.

2 Topic Modeling Overview

A corpus C is a set of documents $\{d_1, \ldots, d_n\}$. A document d_i with $i \in \{1, \ldots, n\}$, is a collection of terms, symbols, and punctuation. After pre-processing, documents are represented as categorical variables where the terms are the categories. A term t_j with $j \in \{1, \ldots, p\}$ is a unique single semantic unit. A topic, denoted by z_m with $m \in \{1, \ldots, k\}$, is a mixture of terms $\{t_1, \ldots, t_p\}$, which is a specific thematic category or subject discussed in C. An assumption in topic modeling is that documents are a mixture of k topics with $m \in \{1, \ldots, k\}$.

2.1 Pre-processing Step

A pre-processing step is usually performed prior to the analysis of text data [7]. This step includes several components: removing stop-words, non-informative terms, and punctuations, defining n-grams, a composition of two or more terms

treated as one, and lemmatization or stemming terms. A document-term matrix, $dtm \in \mathbb{Z}^{n \times p}$, is created indicating the number of terms in each document of the corpus C in its corresponding entry, dtm_{ij} with $i \in \{1, ..., n\}$ and $j \in \{1, ..., p\}$. A step further into the pre-processing step is to define a threshold to remove terms that have high and small importance. The importance of the terms is normally defined through the term-frequency-inverse-document-frequency (tf-idf) matrix, which is based on the relative frequency of the term t_j in the document d_i for all $i \in \{1, ..., n\}$ and $j \in \{1, ..., p\}$, with respect to the total number of terms in d_i, denoted as tf_{ij} and the number of documents where the term t_j appeared, denoted as df_j [14],

$$\omega_{ij} = tf_{ij} \times \log\left(\frac{n}{df_j}\right). \tag{1}$$

2.2 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is one of the most cited methods in topic modeling. It is an unsupervised probabilistic topic modeling Bayesian method introduced in [3]. LDA estimates the probability distribution of topics in documents $\theta = P(z_m|d_i)$ with $i \in \{1, \ldots, n\}$ and $m \in \{1, \ldots, k\}$, and the probability distribution of terms in topic $\phi = P(t_j|z_m)$ with $j \in \{1, \ldots, p\}$. The following priors are considered for the estimation: $\theta \sim \text{Dir}_k(\alpha)$, $\phi \sim \text{Dir}_p(\beta)$, $z_m \sim \text{Mult}(\theta)$ and $t_j \sim \text{Mult}(\phi)$, where Dir(.) represents a symmetric Dirichlet distribution, with α and β serving as hyperparameters that define the sparsity of the topic-document and term-topic relationships, respectively. The estimation of LDA can be achieved through a variational expectation maximization approach or Gibbs sampling.

3 Data Augmentation for Subject Alignment

Generative Pre-trained Transformer (GPT) models such as gpt-3.5-turbo and gpt-4 (developed by OpenAI, https://openai.com/blog/chatgpt#OpenAI) are deep neural network models that employ transformer architecture models pre-trained on extensive textual public datasets, capable of generating both natural language and images.

Data augmentation with ChatGPT is employed to assess subject alignment by generating a set of documents $\{d_1, \ldots, d_n\}$ based on a reference document d_0, which describes the subjects of interest. For efficiency, due to the finite context window of tokens of these models, large documents need to be summarized or pre-processed to extract essential keywords $\{t_{j_{l^*}}, \ldots, t_{j_{f^*}}\}$ with ${j_l}^* \in \{1, \ldots, p\}$ for all $l^* \in \{1, \ldots, f^*\}$, describing the themes in d_0. Although d_0 is a mixture of well-defined k^* topics $\{z_{1^*}, \ldots, z_{k^*}\}$ with $m^* \in \{1^*, \ldots, k^*\}$, these topics are latent and cannot be identified. Typically, topics are statistically identified through replication, which enables the association of co-occurring terms.

To achieve this, ChatGPT is proposed to simulate a sample resembling real-world text. The aim is to generate a set of documents that align with d_0. A

document aligns with d_0 when common topics exist while also covering other subjects. Hence, perfect alignment is not sought but rather partial alignment. Since the topics are unidentified, the subject is defined through keywords. To generate realistic documents integrating topics from d_0 and other topics, a combination of a sample of the keywords of d_0 with a sample of a pre-defined related dictionary $\{t_1, \ldots, t_h\}$, is inputted in ChatGPT for each document. Metadata such as desired output length and context, e.g., informal/formal tone can be provided to guide ChatGPT to generate text based on the aim of the analysis. Other parameters such as the temperature and the top-p threshold (https://platform.openai.com/docs/api-reference/chat/create) can be modified in order to generate less consistent and more diverse text and to validate the process ensuring the authenticity and realistic generation of text documents.

Once the simulated sample of documents $\{d_1, \ldots, d_n\}$ is obtained, an index for subject alignment of the document of interest d_I, needs to be determined. Traditional term frequencies cannot be taken into account since the significance of each keyword varies based on the context in which they are used, impacting their relevance [5,14]. For this reason, the relevance index is computed with the tf-idf (1) for each document d_i where $i \in \{1, \ldots, n\}$ and d_I based on the keywords $\{t_{j_l{}^*}, \ldots, t_{j_f{}^*}\}$, with $j_l{}^* \in \{1, \ldots, p\}$ for all $l^* \in \{1, \ldots, f^*\}$, obtained from d_0,

$$\omega_q = \sum_{j_l{}^*=1}^{p} tf_{qj_l{}^*} \times \log\left(\frac{n+1}{df_{j_l{}^*}}\right), \tag{2}$$

where $q \in \{1, \ldots, n, I\}$.

To assess the subject alignment of a document of interest d_I, various methods can be employed, including visualization through a box-plot and computation of the Z-score: $\frac{\omega_I - \bar{\omega}_{1:n}}{s}$, where $\bar{\omega}_{1:n}$ is the mean of the total relative weights $\{\omega_1, \ldots, \omega_n\}$ and s is the sample standard deviations. A negative Z-score, e.g., usually -2, indicates low alignment between d_I and d_0. The Z-score can be complemented with a hypothesis test for outlier detection such as the Grubbs test.

The generative process for data augmentation for subject alignment of document d_I with d_0 is as follows:

Step 1 Define the following distributions with j_l and $j_l{}^* \in \{1, \ldots, p\}$ for all $l \in \{1, \ldots, f\}$ and $l^* \in \{1, \ldots, f^*\}$: (a) $d_I \sim \text{Mult}(t_{j_l}, \ldots, t_{j_f})$, (b) $d_0 \sim \text{Mult}(t_{j_l{}^*}, \ldots, t_{j_f{}^*})$ and (c) $l_E \sim \text{Mult}(t_1, \ldots, t_h)$
where, (a) and (b) are obtained from the pre-processing step on d_I and d_0 respectively, with probabilities equal to their term relative frequencies, and l_E is a pre-defined list of external related terms with pre-defined probabilities.

Step 2 Sample with replacement a pre-defined number of keywords consisting of r_1 terms from $d_0 \sim \text{Mult}(t_{j_l{}^*}, \ldots, t_{j_f{}^*})$ and r_2 terms from $l_E \sim \text{Mult}(t_1, \ldots, t_h)$.

Step 3 Input the sampled list of keywords in the ChatGPT model along with aim-specific metadata and parameters to generate a variation document representation of d_0.

Step 4 Repeat Step 2 and Step 3 n times to obtain a sample of documents $\{d_1, \ldots, d_n\}$ with $i \in \{1, \ldots, n\}$.

Step 5 Compute the pre-processing step on the generated set of documents $\{d_1, \ldots, d_n\}$ obtained in Step 4 and document of interest d_I to obtain the $dtm \in \mathbb{Z}^{(n+1)\times p}$.

Step 6 Compute the $\{\omega_1, \ldots, \omega_n, \omega_I\}$ weights (2) from the dtm to obtain the representing relevance of each document $\{d_1, \ldots, d_n, d_I\}$ based on the set of keywords $\{t_{j_l*}, \ldots, t_{j_f*}\}$.

Step 7 Compute the Z-score complemented with the hypothesis test for outlier detection, to assess subject alignment of d_I with d_0.

4 Application

4.1 Research Abstracts Alignment with HiTEc's Objectives

The methodology is applied to quantify the alignment of three research abstracts with the scientific objectives of the COST action HiTEc (http://www.hitecaction.org/, Text, functional and other high-dimensional data in econometrics: New models, methods, applications). The proposal of HiTEc is used as the reference document d_0. The three abstracts of interest are defined as d_A, the abstract of this paper, d_B, a philosophical abstract and d_C, an abstract based on digital networking. Both d_B and d_C are non-scientific documents generated with ChatGPT.

The pre-processing step is applied to the HiTEc proposal d_0, to extract in total 189 keywords, including terms like "text", "econometrics", and "computational" and n-grams such as "complex-data" and "high-dimensional". The keywords describe the theme of d_0 and define the distribution $d_0 \sim \text{Mult}(t_{j_l*}, \ldots, t_{j_f*})$ based on their frequencies. The same pre-processing step is applied to abstracts d_A, d_B, and d_C to define their respective distributions, $d_A \sim \text{Mult}(t_{j_l A}, \ldots, t_{j_f A}), d_B \sim \text{Mult}(t_{j_l B}, \ldots, t_{j_f B})$ and $d_C \sim \text{Mult}(t_{j_l C}, \ldots, t_{j_f C})$ respectively.

A total of 1000 abstracts consisting of 150–250 terms are generated with the data augmentation generative process. The parameters are set as 0.2 as temperature and 0.5 as top-p. For each abstract a random sample of 5 terms from the distribution of d_0 and l_E, representing external scientific terms from the CMStatistics conference abstracts, like "autocorrelation", "biostatistics", and "volatility" are integrated into ChatGPT. The goal is to quantify the alignment index for each Corpus: C_A, C_B, and C_C where each Corpus includes the 1000 generated abstracts reflecting the HiTEc action goals along with d_A, d_B, and d_C respectively. The relative total tf-idf (2) weights for d_A, d_B, and d_C are $\omega_A = 0.0013$, $\omega_B = 0$, and $\omega_C = 0.0003$ respectively. LDA is estimated to obtain the term weights within topics on the generated abstracts (Fig. 1). The topic weights within the Corpus are estimated and labeled assign such as "Complex Network": 0.1625, "High-dimensional complex-data": 0.2619, "Distribution analysis": 0.2411, "Biostatistics": 0.1336, "Econometrics": 0.2009. Figure 2 illustrates the box-plot relative total weights for each Corpus. In Corpus C_A, the weight of

d_A exceeds the mean total relative weight $\bar{\omega}_{1:1000}$ of the generated documents, indicating that it aligns with d_0. d_B and d_C as expected have a non-existent to small relative weight. The relative total weight of d_C in C_C comes mainly from the keyword "network", which in d_C appears in the context of digital networking and not as a scientific term.

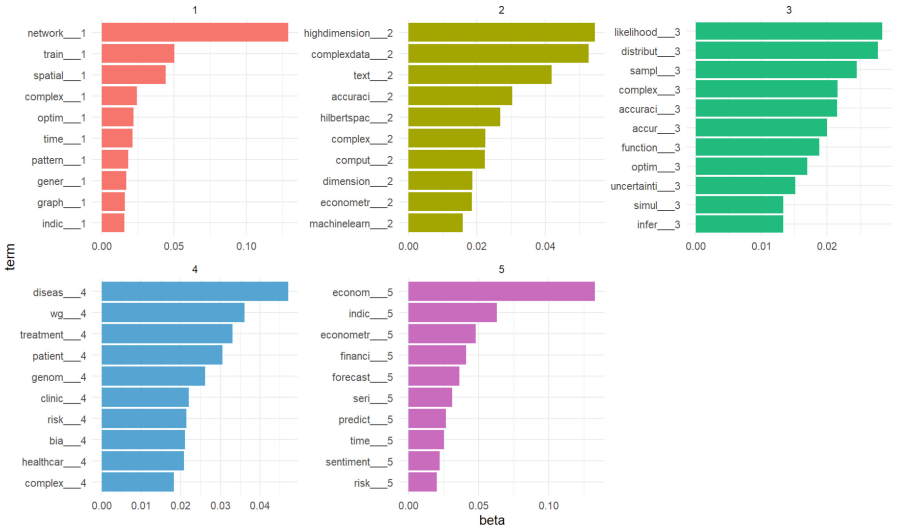


Fig. 1. Terms weights within topics after estimating LDA on the generated abstracts.

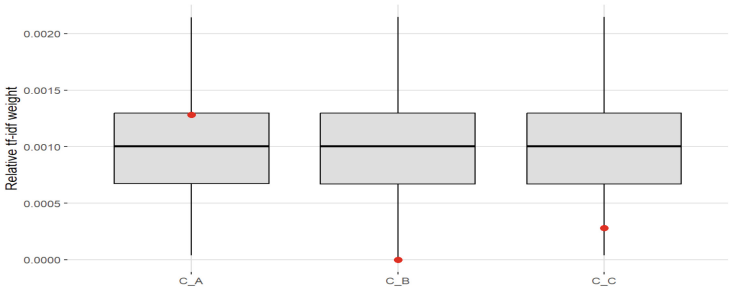


Fig. 2. Box-plots of relative tf-idf weights of C_A, C_B and C_C. In red the corresponding relative weights of d_A, d_B and d_C.

Z-scores are computed to assess how closely documents d_A, d_B, and d_C align with the mean total relative weight $\bar{\omega}_{1:n}$, of the generated documents. The Z-scores of each abstract d_A, d_B and d_C, are found as follows: 0.6736, -2.3642, and -1.7016, respectively, indicating that d_A focuses more on HiTEc compared to d_B and d_C. Despite the Grubbs test indicating no outliers, this result might

stem from the generated abstracts being too conservative, potentially due to the inclusion of too many subjects when they were generated.

4.2 Novel's Chapters Alignment with the Gothic Genre

The methodology is applied to quantify the alignment of the Chapters in Jane Austen's Novel Northanger Abbey with the Gothic genre. ChatGPT is used to generate a dictionary of 364 keywords describing the Gothic genre, such as "terror", "witch", and "darkness" to define the distribution of d_0 and the external list of other genre terms l_E, such as "romance," "military", and "fairy,". A random sample of 55 terms from the distributions d_0 and l_E are obtained for the data augmentation process. In total 800 novel chapters are generated, each containing 1500–2000 terms. Corpus C_{C5} and C_{C6}, representing Chaps. 5 (d_{C5}) and 6 (d_{C6}) of Northanger Abbey, along with the 800 generated chapters, are analysed to quantify alignment with the Gothic genre. The relative total tf-idf weights (2) for d_{C5} and d_{C6} are $\omega_{C5} = 0$ and $\omega_{C6} = 0.0028$ respectively, with corresponding Z-scores of -3.4626 and 11.9142, indicating that Chap. 6 has a high index weight of the Gothic genre in comparison to Chap. 5. No outliers were indicated when the Grubbs test was conducted.

5 Conclusion and Future Research

An innovative data augmentation method using ChatGPT is proposed to evaluate subject alignment when text data are scarce. This approach involves evaluating subject alignment by comparing text with a reference document using keywords addressing the identification problem. By incorporating the keywords and external related terms, a sample is generated to introduce variability and meaningful noise. ChatGPT, combined with tf-idf and Z-score index, evaluates subject alignment, complemented with a hypothesis test for outlier detection. To improve the quality of the generated documents, fine-tuning the GPT model with domain-specific text datasets can be implemented to effectively train the model.

Acknowledgements. The (financial) collaboration provided by the COST Action HiTEc CA21163 and the Qatar Centre for Global Banking and Finance is acknowledged.

References

1. Anaby-Tavor, A., et al.: Do not have enough data? Deep learning to the rescue!. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34(05), pp. 7383–7390. (2020)
2. Bird, J.J., Faria, D.R., Ekárt, A., Premebida, C., Ayrosa, P.P.S.: LSTM and GPT-2 Synthetic Speech Transfer Learning for Speaker Recognition to Overcome Data Scarcity.arXiv:2007.00659 (2020)
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)

4. Chuang, J., Gupta, S., Manning, C., Heer, J.: Topic model diagnostics: assessing domain relevance via topical alignment. In: International Conference on Machine Learning, pp. 612–620. PMLR (2013)
5. Elhadad, N., Barzilay, R.: Sentence alignment for monolingual comparable corpora. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pp. 25–32. (2003)
6. Fernandez, V., et al.: Can segmentation models be trained with fully synthetically generated Data? In: Zhao, C., Svoboda, D., Wolterink, J.M., Escobar, M. (eds) Simulation and Synthesis in Medical Imaging. SASHIMI 2022. Lecture Notes in Computer Science, vol. 13570, pp. 79–90. Springer, Cham (2022)
7. Gentzkow, M., Kelly, B., Taddy, M.: Text as data. J. Econ. Lit. **57**(3), 535–574 (2019)
8. Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., Kapadia, S.: Making text count: economic forecasting using newspaper text. J. Appl. Econom. **37**(5), 896–919 (2022)
9. Kontoghiorghes, L., Colubi, A.: Testing the homogeneity of topic distribution between documents of a Corpus. In: García-Escudero, L.A., et al. Building Bridges between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing, vol. 1433, pp. 248–254. Springer, Cham (2023)
10. Kontoghiorghes, L., Colubi, A.: New metrics and tests for subject prevalence in documents based on topic modeling. Int. J. Approx. Reason. **157**, 49–69 (2023)
11. Patel, N., Nagpal, P., Shah, T., Sharma, A., Malvi, S., Lomas, D.: Improving mathematics assessment readability: do large language models help? J. Comput. Assist. Learn. **39**(3), 804–822 (2023)
12. Prezja, F., Annala, L., Kiiskinen, S., Ojala, T.: Exploring the efficacy of base data augmentation methods in deep learning-based radiograph classification of knee joint osteoarthritis. Algorithms **17**(1), 8 (2024)
13. Reddy, D., Marasani, J., Sathwik, P., Gowtham, V., Abhishek, S.G., Anjali: an empirical analysis of topic categorization using PaLM, GPT and BERT models. In: 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1–6. (2023)
14. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. **24**(5), 513–523 (1988)
15. Sun, X., He, J.: A novel approach to generate a large scale of supervised data for short text sentiment analysis. Multimed. Tools Appl. **79**(9), 5439–5459 (2020)
16. Yoo, K.M., Park, D., Kang, J., Lee, S.-W., Park, W.: GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation (2021). arXiv:2104.08826
17. Zhang, M., Naresh, N.U., Yuxiong He, Y.: Adversarial data augmentation for task-specific knowledge distillation of pre-trained transformers. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36(10), pp. 11685–11693. (2022)

The Impact of Sentiment in Social Network Communication

Soveatin Kuntur and Anna Cena(✉)

Faculty of Mathematics and Information ScienceWarsaw, University of Technology, Koszykowa 75, 00-662 Warsaw, Poland
anna.cena@pw.edu.pl

Abstract. The pervasive influence of social media in shaping public discourse and informing societal and policy matters cannot be denied. Several attempts were made in order to investigate the nature/dynamics of this process, both, from social networks and sentiment analysis perspective. Few more has set a goal of combining aforementioned approaches. Contribution of this paper continues further in this direction. We will take a closer look into the network specific measures that allow to aggregate the impact of a given user. The main focus of this contribution, however, is to investigate the interaction of content with varied polarities on to the social network structure itself, and the final assessment of each user's importance. We analyzed a Twitter dataset, including various contentious socio-political topics, assessing the structure of engagement networks (including retweets, mentions, etc.) and the sentiment of the tweets. We explored multiple methods of integrating sentiment into the network, such as considering it as a weight factor for connections.

Keywords: Sentiment analysis · Social networks · Importance · Data aggregation

1 Introduction

The use of various social media platforms to analyze social dynamics has been increasingly popular in recent years. Amongst them Twitter can be distinguished as a prime example. Twitter not only illuminates the connections among its users, but also provides an in-depth view into their perspectives and prevailing sentiment. Moreover, in today's context, it has evolved beyond a social media platform for opinion exchange, acting as a news source for thousands of individuals. Users publish their content in real time, reacting very quickly to social, economic and political events. There are many points of view that may be of interest while analyzing this kind of data. For example, natural language processing methods can be incorporated in order to detect spammers and fake news (cf. e.g. [2,18]), while network approach allows to investigate the relationships and dependencies

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 225–233, 2024.
https://doi.org/10.1007/978-3-031-65993-5_27

within the users community (cf. e.g. [6]). The focus of this paper, however, lies predominantly on identifying users who can be considered as significant or influential within the network structure, with their content potentially shaping the tenor of the discourse. In past years, several studies have embarked on a similar quest. For instance, the study conducted by [19] denotes influential users as those whose content has the capacity to gain extensive reception and dissemination. In order to detect them the structural properties of the retweet network are utilized. The influence of a user is measured by their in-degree (number of retweets) and the number of unique tweets. Furthermore, the paper [19] also calculates the cumulative influence of all users in a community to identify influential communities. On the other hand, in [5] the influence of Twitter users was estimated using their retweet h-index, which is an adaptation of the famous Hirsch index specifically to Twitter. The retweet h-index takes into account the number of tweets posted, as well as the impact of individual tweets in terms of retweets. In both cases [5,19] the information carried out by the retweets is used in order to construct a network of users, represented as a directed graph. However, in contrast to the aforementioned papers, the research conducted by [3] employs retweets to construct an undirected graph and utilizes a variety of techniques to detect significant users with the information regarding the time in which they post/react, the number of followers, the variety of the topics they contribute to, expertise etc. Another interesting approach incorporate various Twitter metrics that allows to indicate the importance of a given user (cf. e.g. [13,17,22]). One may also find publications that propose PageRank-like algorithms to do the job. Various modification of original approach ([15]) were proposed in order to recursively compute users' influence (see e.g. [10,20,21]).

Please note that even though we live in undeniable era of information explosion, the access to many data sources become (for various reasons) more and more restricted (e.g. social media platforms do not allow for data scraping or fees are introduced in order to do so). Therefore, we will explore what knowledge can be found using the set of data that is available. In this paper we have adopted the information fusion approach—we decided to focus on social network analysis with the sentiment of each posts included as an additional factor. The main idea is to be able to not only analyse the significance of each user but also the overall sentiment of the message they spread.

The paper is organized as follows. In Sect. 2 the overall of the methodology is presented, with the short theoretical description of the graph-based influence measures and chosen sentiment analysis algorithms. Next, Sect. 3 cover the details of network structure as well as present the results of our empirical analysis. Finally, Sect. 4 summarizes and concludes the contribution.

2 Capturing User's Influence

Let's assume that the network of interest is represented as directed, weighted graph $G = (V, E)$, where $V(G) = \{v_1, v_2, \ldots, v_n\}$ denotes the set of all possible vertices and $E(G) = \{e_{uv} : u, v \in V(G)$ such that there is as edge from u to $v\}$

represent the edges. In our case there is an edge from user u to user v if u interacted with users' v post in any way (retweet, mention, replay). Additionally, the $w(e_{uv})$ denotes the weight of such connection. Typically we assume the number of aforementioned interactions to be the weight of a connection (cf. Sect. 3.2).

In the following section we will introduce some graph-based metrics that allows us to indicate the users importance based solely on the number of connection he/she has.

2.1 Graph Based User's Influence Measures

Centrality measures are among the most widely used indices based on network data. They generally reflect a unit's prominence (e.g. status, prestige or visibility (see [12])). Such measures allow to capture how "central" a node is in the network, which can be used as an estimate of its overall importance (see [8]). In order to measure influence one can include following graph-based metrics (see [7]).

Node degree. Node degree (also called degree-centrality) is simply defined as the number of edges (connections) node u has in the graph. The interpretation is very simple: the higher the degree, the more central the node is. Of course for directed graphs one should include both in-degree, i.e., $C_{in}(u) = |\{v \in V(G) : \text{ such that } e_{vu} \in E(G)\}|$, and out-degree, i.e., $C_{out}(u) = |\{v \in V(G) : \text{ such that } e_{uv} \in E(G)\}|$.

Betweenness centrality. Betweenness centrality is given by $C_B(u) = \sum_{w,v \in V(G)} \frac{p_{v,w}(u)}{p_{v,w}}$, where $p_{v,w}$ denotes the number of geodesics (i.e., sequence of vertices connected by edges between vertices v and w with the fewest possible edges between) and $p_{v,w}(u)$ is the number of geodesics between node v and w that pass through u. This measure characterizes the importance of a given node for establishing short pathways between other nodes.

Eigenvector centrality. Eigenvector centrality of a node is based on the centrality of its neighbors. The eigenvector centrality for node v_i is the i-th element of the vector $\mathbf{x}$ defined by the equation $A\mathbf{x} = \lambda\mathbf{x}$, where A is the adjacency matrix of the graph G with eigenvalue λ.

Closeness centrality. Closeness centrality indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network, i.e.: $C(u) = \frac{n-1}{\sum_{v \neq u} d(v,u)}$, where $d(v,u)$ is the number of edges in the geodesic linking u and v.

PageRank-based algorithms. PageRank-like algorithms are based on the analysis of a graph structure to discover the impactful nodes (users in our case) and often such solutions outperform other methods. Note that PageRank-based solutions take into account the global link structure and weigh graph nodes when measuring influence in social networks. For example, topic-sensitive influence is measured with the usage of TwitterRank [20]. On the other hand, users authority is estimated with the help of TuRank [21] algorithm.

2.2 Sentiment Driven Influence Analysis

Sentiment analysis, one of Natural Language Processing (NLP) tasks, aims to understand people's emotions, attitudes, and opinions from the written language [14]. Sentiment analysis can provide invaluable insights across various sectors, such as marketing, politics, customer service, and product development. Researchers have explored numerous methods, with machine learning models like RoBERTa [11] and lexical approaches like VADER [9] becoming increasingly popular.

We are going to use two different (both in concept and application) approaches in order to extract the information regarding sentiment of tweets.

First, we apply RoBERTa algorithm, introduced by [11]. This is a variation of the transformer-based model BERT, optimized for more robust performance on NLP tasks. It removes the next sentence prediction objective and dynamically adjusts the masking pattern applied to the model's input [11]. The model's robustness is particularly effective for sentiment analysis, allowing for comprehensive understanding of semantic meanings, including negation and intensification [1]. Empirical evaluations have shown that RoBERTa outperforms many previous *state-of-the-art* models on several sentiment analysis benchmarks [23].

We also decided to compare results with the ones obtained via VADER (Valence Aware Dictionary for Sentiment Reasoning)—a rule-based sentiment analysis tool incorporated into the Natural Language Toolkit (NLTK) [9]. Unlike models like RoBERTa, which require substantial computational resources, VADER uses a combination of lexical heuristics and grammatical rules, making it computationally efficient [9]. VADER is specifically designed to handle social media text, including the use of slangs, acronyms, and emoticons. Despite its simplicity, VADER performs remarkably well on standard sentiment analysis datasets and competes with more complex models [16].

3 Empirical Analysis

3.1 Sentiment Specific Networks

Firstly, we decided to split our network into three separate parts, build with only neutral, positive and negative tweets, respectively. This classification was obtained via RoBERTa algorithm. Please note that around 56% of all considered tweet were labeled as neutral, 37% as positive and only 7% as negative. Table 1 presents the percentage of common users between networks. Each of considered networks was build as a sub-graph of the original structure that includes only those users (represented as nodes) and connections that refer to specific tweet group. Interestingly, the proportion of users included in each network corresponds roughly to the distribution of sentiment classes. Moreover, we may observe that the number of common users between those graphs was not greater than 10%. This would suggest that specific users produce post of which majority can fall into one of sentiment categories.

Table 1. The common users distribution between networks in %.

	Neutral	Positive	Negative
Whole network	55.43	48.91	6.53
Neutral	100	9.16	1.86
Positive	9.16	100	3.15
Negative	1.86	3.15	100

The rankings of importance for each node were calculated based on centrality measures (degree, eigenvector, clossness, betweenes), Page Rank output as well as weighted in- and out-degree values. The correlation between them was measured via Spearman ρ correlation coefficient. Note that this statistic does not depend on the raw values but only on their ranks. Figure 1a depicts the correlation heatmap created for the whole network, however, similar behavior can be observed for sentiment-specific network as well. Considered measures present positive but very week correlation with the exception of out-degree values—here quite strong negative correlation may be observed. In particular case of weighted in- and out-degrees, the -0.8 indicate that users can be divided into content creators (i.e. the ones who produce the most tweets) and content distributors (i.e. users that mostly retweet).

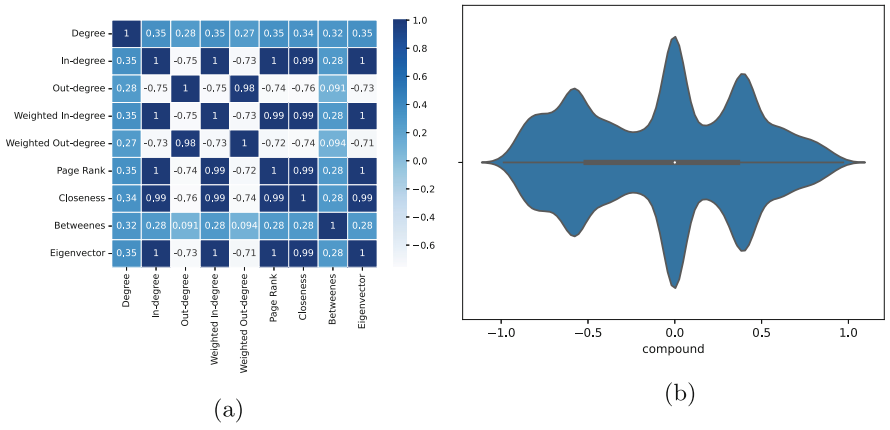


Fig. 1. (a) Spearman's ρ calculated for each influence measures considered. (b) Violin plot of the compound (sentiment degree) distribution obtained via VADER algorithm.

In order to assess importance of each user, we decided to combine all information by using the Borda Count (see e.g. [4]) method of rankings aggregation. Rankings created with each of 8 centrality and importance approaches where then treated as "ballots" . For each ballot, a number of points corresponding to the number of options ranked lower is assign and the one with the most points is

Table 2. Top 5 users according to the Borda Count Aggregation with influence measures.

Network	User	d_{cent}	e_{cent}	C	B⋆	PR	$d_{w,in}$	$d_{w,out}$
All	RpsAgainstTrump	328	0.0000	0.0145	0.0000	0.0065	325	5
	anti_putin_bot	20	0.5874	0.0004	0.0000	0.0004	19	27
	kermoost	13	0.1021	0.0003	0.0000	0.0003	9	8
	jacobnhls	15	0.0000	0.0005	0.0000	0.0002	12	3
	donald_j_putin	7	0.0001	0.0002	0.0000	0.0002	5	8
Neutral	anti_putin_bot	13	0.6120	0.0004	0.0000	0.0005	7	15
	PatWay1979	9	0.0000	0.0006	0.0000	0.0003	10	3
	jacobnhls	10	0.0000	0.0006	0.0000	0.0003	7	3
	dv_devain	4	0.0000	0.0003	0.0000	0.0004	8	4
	PaulaChertok	21	0.0000	0.0016	0.0000	0.0007	22	2
Positive	anti_putin_bot	11	0.5902	0.0004	0.0000	0.0004	10	10
	addison_dr	11	0.0000	0.0004	0.0000	0.0002	5	7
	carlquintanilla	18	0.0000	0.0015	0.0000	0.0006	16	2
	donald_j_putin	6	0.0689	0.0003	0.0000	0.0004	3	3
	RpsAgainstTrump	6	0.0000	0.0003	0.0000	0.0002	3	3
Negative	kermoost	11	0.6742	0.0034	0.0000	0.0053	5	7
	FreeUkraine91	19	0.0002	0.0110	0.0000	0.0091	17	2
	MeNescire	6	0.0000	0.0027	0.0000	0.0017	4	2
	twolilsZ	4	0.0000	0.0014	0.0000	0.0008	2	7
	anti_putin_bot	3	0.0000	0.0014	0.0000	0.0010	2	2

Table 3. Top 5 users according to the Borda Count Aggregation of influence measures based on sentiment weighted network.

User	d_{cent}	e_{cent}	C	B⋆	PR	$d_{w,in}$	$d_{w,out}$
anti_putin_bot	20.0000	0.5874	0.0004	0.0000	0.0003	8.1329	9.4857
FreeUkraine91	44.0000	0.0000	0.0018	0.0000	0.0008	22.2687	1.6035
PaulaChertok	22.0000	0.0000	0.0009	0.0000	0.0003	19.6539	1.8689
RpsAgainstTrump	328.0000	0.0000	0.0145	0.0000	0.0068	94.7376	1.3817
kermoost	13.0000	0.1021	0.0003	0.0000	0.0003	5.3435	3.9931

the winner. Table 2 presents the values of considered measures for top 5 users in each network. Please note that betweenes are very small (due to rounding they appear to be equal 0).

Table 4. Number of tweets, retweets and average retweet count for top 5 users according to the Borda Count Aggregation of influence measures based on sentiment weighted network.

	User	Number of Tweets	Number of Retweets	Average Retweet
1	anti_putin_bot	29	19	0.7
2	FreeUkraine91	3	50	16.7
3	PaulaChertok	2	27	13.5
4	RpsAgainstTrump	5	325	65.0
5	kermoost	8	9	1.1

3.2 Sentiment Weighted Network

Our second approach is based on incorporating weights directly into the network structure. In this case we decided to use VADER algorithm, which instead of final classification of posts, returns degree of sentiment (i.e. values between -1 and 1, where -1 denotes extreme negative sentiment and 1—extreme positive sentiment). The distribution of returned sentiment degree is presented on Fig. 1b. Please note those values seems to follow tri-modal and quite symmetric distribution. It is important to note that we decided to incorporate absolute value of this compound in order to capture the strength of emotion rather then its direction. Each edge between two users was then weighted not by the number of retweets but as the sum of their sentiments. Please note that since we included all kinds of possible interactions the compound used as a weight refers to either original post or to the replay tweet. Table 3 present the top 5 users detected within sentiment weighted network together with their influence characteristics. Please note that detected users are consistent within those found in sentiment specific networks. Moreover, Table 4 present the quantity measures regarding given user activity.

4 Conclusions

The investigation carried out in this paper focuses on information flow within the social network on the example of Tweeter platform. Taking the approach of combining complex networks tools and sentiment-based analysis of tweets themselves. We aimed at the influencial users detection. Similar approaches were conducted before that mostly focus on complex-network modelling of the retweet network. We, however, decided to include the sentiment of the content into the network structure itself. Moreover, we defined connection between two users not only by the means of retweet but also other types of interactions, i.e., mentions, replays. In order to detect important users we created rankings based of commonly-used structural properties of the engagement network (i.e. various centrality measures, page rank, weighted in- and out-degrees). Final ranking was obtained with the means of Borda Count voting aggregation. The Borda Count

is intended to select broadly acceptable option, rather than those preferred by a majority. Therefor, it can be viewd as a consensus-based voting system. Our first approach was to divide the network into sub-graphs containing only tweets of specific sentiment. In order to do so, we use RoBERTa algorithm in order to classify tweets into three separate categories: neutral, positive and negative. We observe that users within each sentiment-based network differ, with the highest percentage of common users being less the 10%. This suggest that sentiment of statements is rather consistent within the specific users. Note however, that most of the tweets (around 56%) were classify as neutral. In order to verify evaluate those results, we use VADER algorithm in order to obtain the sentiment compound for each tweet, instead of the crisp grouping. The compound coefficient (that ranges form -1 to 1, i.e. extreme negative to extreme positive sentiment) was used then as the weight of edge representing the interactions between users. Interestingly, the top 5 users detected this way consist with the set obtained across our sentiment-based networks.

The preliminary results presented in this paper depend heavily on the choice on algorithm for sentiment detection. We compare two different approaches, but we believe that the comparison study on larger dataset is needed. Moreover, some additional factors could be included as well.

References

1. Dai, J., Yan, H., Sun, T., Liu, P., Qiu, X.: Does syntax matter? A strong baseline for aspect-based sentiment analysis with roberta. arXiv preprint arXiv:2104.04986 (2021)
2. Devarajan, G.G., Nagarajan, S.M., Amanullah, S.I., Mary, S.A.S.A., Bashir, A.K.: AI-assisted deep NLP-based approach for prediction of fake news from social media users. IEEE Trans. Comput. Soc. Syst. 1–11 (2023)
3. Eliacik, A.B., Erdogan, N.: Influential user weighted sentiment analysis on topic based microblogging community. Expert Syst. Appl. **92**, 403–418 (2018)
4. Emerson, P.: The original Borda count and partial voting. Soc. Choice Welf. **40**, 353–358 (2013)
5. Evkoski, B., Pelicon, A., Mozetič, I., Ljubešić, N., Kralj Novak, P.: Retweet communities reveal the main sources of hate speech. PLoS ONE **17**(3), e0265602 (2022)
6. Feng, X., Zhao, Q., Zhu, R.: Towards popularity prediction of information cascades via degree distribution and deep neural networks. J. Informetr. **17**(3), 101413 (2023)
7. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)
8. Golbeck, J.: Chapter 3—Network structure and measures. In: Golbeck, J. (ed.) Analyzing the Social Web, pp. 25–44. Morgan Kaufmann, Boston (2013)
9. Hutto, C., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, pp. 216–225 (2014)
10. Li, X., Cheng, S., Chen, W., Jiang, F.: Novel user influence measurement based on user interaction in microblog. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13, pp. 615–619. Association for Computing Machinery, New York, NY, USA (2013)

11. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
12. Marsden, P.V.: Network analysis. In: Kempf-Leonard, K. (ed.) Encyclopedia of Social Measurement, pp. 819–825. Elsevier, New York (2005)
13. Muñoz, M.M., de Gracia, M.M.R., Navas-Sarasola, C.: Measuring engagement on twitter using a composite index: an application to social media influencers. J. Informetr. **16**(4), 101323 (2022)
14. Nasukawa, T., Yi, J.: Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP'03, pp. 70–77. Association for Computing Machinery, New York, NY, USA (2003)
15. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. Technical Report 1999–66, Stanford InfoLab (1999)
16. Ribeiro, F.N., Araújo, M., Gonçalves, P., André Gonçalves, M., Benevenuto, F.: SentiBench—a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci. **5**, 1–29 (2016)
17. Riquelme, F., González-Cantergiani, P.: Measuring user influence on twitter: a survey. Inform. Process. Manag. **52**(5), 949–975 (2016)
18. Shahzad, M., Alhoori, H., Freedman, R., Rahman, S.A.: Quantifying the online long-term interest in research. J. Informetr. **16**(2), 101288 (2022)
19. Sluban, B., Smailović, J., Battiston, S., Mozetič, I.: Sentiment leaning of influential communities in social networks. Comput. Soc. Netw. **2**, 9 (2015)
20. Weng, J., Lim, E.-P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10, pp. 261–270. Association for Computing Machinery (2010)
21. Yamaguchi, Y., Takahashi, T., Amagasa, T., Kitagawa, H.: TURank: Twitter user ranking based on user-tweet graph analysis. In: Chen, L., Triantafillou, P., Suel, T. (eds.) Web Information Systems Engineering—WISE 2010. Lecture Notes in Computer Science, vol. 6488, pp. 240–253. Springer, Berlin, Heidelberg (2010)
22. Zhang, M., Zhang, D., Zhang, Y., Yeager, K., Fields, T.N.: An exploratory study of twitter metrics for measuring user influence. J. Informetr. **17**(4), 101454 (2023)
23. Zhang, T., Xu, B., Thung, F., Haryono, S.A., Lo, D., Jiang, L.: Sentiment analysis for software engineering: How far can pre-trained transformer models go? In: 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 70–80. IEEE (2020)

Constructing Measures of Dependence Via Sensitivity of Conditional Distributions

Patrick B. Langthaler(✉), Jonathan Ansari, Sebastian Fuchs, and Wolfgang Trutschnig

Department of Artificial Intelligence and Human , InterfacesParis-Lodron University Salzburg, Salzburg, Austria
{patrick.langthaler,jonathan.ansari,sebastian.fuchs}@plus.ac.at, wolfgang@trutschnig.net

Abstract. An essential objective in statistics is to determine the relationship, specifically the dependence, between two random variables X and Y. To this end many classic and novel methods have been described. Here we show that some of these methods can be seen as special cases of a broad class of measures of dependence, which essentially measure the sensitivity of the distribution of Y conditional on X to changes in X. We introduce some new members of this class which characterise dependence in the sense that they are equal to 0 if and only if X and Y are independent and are equal to 1 if and only Y is a measurable function of X. We establish some further useful properties of these measures and discuss their estimation. In particular, for continuous (X, Y), the introduced measures of dependence rely only on the underlying copula of (X, Y) and not on the marginal distributions. This allows the use of Checkerboard Copulas to obtain a strongly consistent estimator.

Keywords: Measures of dependence · Sensitivity · Conditional distributions · Checkerboard copulas

Note The following is a summary of the results of [2] where more results, their proofs as well as a simulation study and a real data example can be found.

1 Motivation and Introduction

1.1 Notation and General Setting

Throughout this contribution let $(X, Y), (X_1, Y_1), ..., (X_n, Y_n)$ be an i.i.d sample of some bivariate distribution function. We denote the measure induced by the distribution function of X by $\mathbb{P}^X$, the measure induced by the distribution function of Y by $\mathbb{P}^Y$ and the measure induced by the distribution function of Y conditional on $X = x$ as $\mathbb{P}^{Y|X=x}$. We furthermore denote the distribution function of Y as F_Y and the distribution function of Y conditional on $X = x$ as $F_{Y|X=x}$.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 234–240, 2024.
https://doi.org/10.1007/978-3-031-65993-5_28

1.2 Constructing Measures of Dependence

It is a common statistical procedure to characterize the dependence between a binary random variable X and another random variable Y by comparing $F_{Y|X=0}$ with $F_{Y|X=1}$, that is computing some

$$\Delta(\mathbb{P}^{Y|X=0}, \mathbb{P}^{Y|X=1})$$

for example

$$\Delta(\mathbb{P}^{Y|X=0}, \mathbb{P}^{Y|X=1})$$
$$= \begin{cases} \sup_y |F_{Y|X=0}(y) - F_{Y|X=1}(y)| & \text{Kolmogorov-Smirnov} \\ \int_{\mathbb{R}} (F_{Y|X=0}(y) - F_{Y|X=1}(y))^2 d\mathbb{P}^{Y|X=1}(y) & \text{Cramer-von-Mises} \\ \dfrac{\mathbb{E}(Y|X=0) - \mathbb{E}(Y|X=1)}{\sqrt{\mathbb{P}(X=0)\mathbb{V}(Y|X=0) + \mathbb{P}(X=1)\mathbb{V}(Y|X=1)}} & \text{Standardised Mean Difference} \end{cases}$$

A sensible extension to a discrete variable X which can take more than two values (say k different values) is then the sum (or mean) over all possible pairwise comparisons

$$\sum_{i,j=1}^{k} \Delta(\mathbb{P}^{Y|X=x_i}, \mathbb{P}^{Y|X=x_j}).$$

since this gives an idea of how strongly the distribution of Y varies when changing X. For general distributions (X, Y) and after weighting by the measure of X this then becomes the integral

$$\int_{\mathbb{R}^2} \Delta(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}) d(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2).$$

For purposes of normalisation it is advisable to divide by

$$\alpha_\Delta := \sup \left\{ \int_{\mathbb{R}^2} \Delta(\mathbb{P}^{Y'|X'=x_1}, \mathbb{P}^{Y'|X'=x_2}) d(\mathbb{P}^{X'} \otimes \mathbb{P}^{X'})(x_1, x_2) | (X', Y') \in \mathcal{R}(Y) \right\}$$

where $\mathcal{R}(Y)$ is the class of all random vectors (X', Y') such that $Y' \stackrel{d}{=} Y$. Thus we define

$$\Lambda_\Delta(Y|X) := \alpha_\Delta^{-1} \int_{\mathbb{R}^2} \Delta(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}) d(\mathbb{P}^X \otimes \mathbb{P}^X)(x_1, x_2)$$

It turns out that several well established classical as well as novel measures of dependence can be written in this fashion.

Example 1. Defining

$$\Delta(\mathbb{P}^{Y|X=0}, \mathbb{P}^{Y|X=1}) := (\mathbb{E}(Y|X=x_1) - \mathbb{E}(Y|X=x_2))^2$$

leads to

$$\alpha_\Delta = 2\mathbb{V}(Y)$$

and

$$\Lambda_\Delta(Y|X) := \frac{\mathbb{V}(\mathbb{E}(Y|X))}{\mathbb{V}(Y)}$$

which is the proportion of the variance of Y explained by regressing on X, sometimes also referred to as the Sobol index or the Cramer-von-Mises index (see [10]).

Example 2. Defining

$$\Delta(\mathbb{P}^{Y|X=0}, \mathbb{P}^{Y|X=1}) = \int_{\mathbb{R}} (\mathbb{P}(Y \geq y|X = x_1) - \mathbb{P}(Y \geq y|X = x_2))^2 d\mathbb{P}^Y(y)$$

leads to

$$\alpha_\Delta = \int_{\mathbb{R}} \mathbb{V}(\mathbb{1}_{\{Y \geq y\}}) d\mathbb{P}^Y(y)$$

and

$$\Lambda_\Delta(Y|X) := \frac{\int_{\mathbb{R}} \mathbb{V}(\mathbb{P}(Y \geq y|X)) d\mathbb{P}^Y(y)}{\int_{\mathbb{R}} \mathbb{V}(\mathbb{1}_{\{Y \geq y\}}) d\mathbb{P}^Y(y)}$$

which is, of course, Chatterjee's famous coefficient of correlation [5].

Comparing conditional distributions is not a novel idea, but to our knowledge most methods compare the conditional with the unconditional distribution of Y, that is

$$\Lambda_\Delta(Y|X) := \alpha_\Delta^{-1} \int_{\mathbb{R}} \Delta(\mathbb{P}^{Y|X=x}, \mathbb{P}^Y) d\mathbb{P}^X(x)$$

see for example [1,3,5,7,9,10,12,14,15,19,20,22–24]. It can be shown that in some cases, for example in the previous example both notions coincide. However this is not the case in general.

For the remainder of this contribution we will focus on measures of dependence induced by

$$\Delta_\varphi(\mathbb{P}^{Y|X=x_1}, \mathbb{P}^{Y|X=x_2}) = \int_{\mathbb{R}} \varphi(F_{Y|X=x_1}(y) - F_{Y|X=x_2}(y)) d\mathbb{P}^Y(y)$$

and will identify

$$\Lambda_\varphi(Y|X) := \Lambda_{\Delta_\varphi}(Y|X)$$

where $\varphi : [-1, 1] \longrightarrow \mathbb{R}$ is convex and strictly convex in zero with $\varphi(0) = 0$. Note that as per the example above, Chatterjee's Rank Correlation is a special

case of this class of measure with $\varphi(x) = x^2$. It can be shown that for this class of dependence measure the normalising constant is

$$\alpha_\varphi := \alpha_{\Delta_\varphi} = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}) d\mathbb{P}^Y(y) d(\mathbb{P}^Y \otimes \mathbb{P}^Y)(y_1, y_2)$$
$$= \frac{1}{6}\varphi(-1) + \frac{1}{6}\varphi(1).$$

2 Properties

It is clear from the definition and due to the normalisation that $\Lambda_\varphi(Y|X) \in [0,1]$. The most important property, which justifies the use of the term *measure of dependence* (see e.g. [1,4–6,13,21,23]), is the following

Theorem 1. *$\Lambda_\varphi(Y|X)$ as introduced above fulfills the following properties:*

1. *$\Lambda_\varphi(Y|X) = 0$ if and only if X and Y are independent.*
2. *$\Lambda_\varphi(Y|X) = 1$ if and only if Y is completely dependent on X, i.e. there exists a measureable function g such that $Y = g(X)$.*

Since our objective is to quantify the dependence of Y on X irrespective of any underlying distributional assumptions a desirable property would be for $\Lambda_\varphi(Y|X)$ to be invariant to e.g. changes in units and/or logarithmising or exponentiating Y. Indeed we have the following

Proposition 1. *$\Lambda_\varphi(Y|X)$ fulfills the following invariance properties:*

1. *$\Lambda_\varphi(Y|X) = \Lambda_\varphi(Y|h(X))$ for all bijective $h : \mathbb{R} \longrightarrow \mathbb{R}$*
2. *$\Lambda_\varphi(Y|X) = \Lambda_\varphi(g(Y)|X)$ for all strictly increasing $g : \mathbb{R} \longrightarrow \mathbb{R}$*

The functional $\Lambda_\varphi(Y|X)$ also fulfills the so-called *Data Processing inequality* (see [21]) and *Self-Equitability* (see [8,17]) which, roughly speaking, ensure that

1. the value of $\Lambda_\varphi(Y|X)$ cannot be increased by some convenient transformation of X.
2. transformations of X that are in a sense *sufficient* to describe the relationship between Y and X leave $\Lambda_\varphi(Y|X)$ unchanged.

More precisely we have:

Proposition 2. *For all measurable $h : \mathbb{R} \longrightarrow \mathbb{R}$*

$$\Lambda_\varphi(Y|h(X)) \leq \Lambda_\varphi(Y|X)$$

with equality when Y and X are independent conditional on $h(X)$.

It is also desirable that when the distributions of (X,Y) and some (X',Y') are *similar*, the values of $\Lambda_\varphi(Y|X)$ and $\Lambda_\varphi(Y'|X')$ are also close. This is ensured by

Proposition 3. *Let $(X,Y),(X_1,Y_1),(X_2,Y_2),...$ be bivariate random vectors with continuous distributions functions and corresponding copulas $A, A_1, A_2, ...$ Let φ be continuous. Then weak conditional convergence (as defined in [16]) of A_n to A implies convergence of $(\Lambda_\varphi(Y_n|X_n))_{n\in\mathbb{N}}$ to $\Lambda_\varphi(Y|X)$.*

3 Estimation

After establishing the useful properties of the functional $\Lambda_\varphi(Y|X)$ the immediate question that arises, especially for the practitioner, is how to actually estimate this quantity, given a set of data. When X is discrete, and plentiful observations for each value X can take are available, $F_{Y|X=x_i}(y)$ may be estimated with the usual methods and pluging these estimates into $\Lambda_\varphi(Y|X)$ gives a straightforward plug-in estimator. In the continuous setting however, while a plug-in estimator can in principle be constructed, $\hat{F}_{Y|X=x_i}(y)$ will simply be $\mathbb{1}_{\{y_i \leq y\}}$. Thus, for a given dataset $(x_1, y_1), ..., (x_n, y_n)$ drawn from a continuous bivariate distribution, the denominator of a plug-in estimator of $\Lambda_\varphi(Y|X)$ becomes

$$\begin{aligned}
&\int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(\hat{F}_{Y|X=x_1}(y) - \hat{F}_{Y|X=x_2}(y)) d\hat{\mathbb{P}}^Y d(\hat{\mathbb{P}}^X \otimes \hat{\mathbb{P}}^X)(x_1, x_2) \\
&= \frac{1}{n^3} \sum_{i,j,k=1}^{n} \varphi(\hat{F}_{Y|X=x_i}(y_k) - \hat{F}_{Y|X=x_j}(y_k)) \\
&= \frac{1}{n^3} \sum_{i,j,k=1}^{n} \varphi(\mathbb{1}_{\{y_i \leq y_k\}} - \mathbb{1}_{\{y_j \leq y_k\}}) \\
&= \int_{\mathbb{R}^2} \int_{\mathbb{R}} \varphi(\mathbb{1}_{\{y_1 \leq y\}} - \mathbb{1}_{\{y_2 \leq y\}}) d\hat{\mathbb{P}}^Y(y) d(\hat{\mathbb{P}}^Y \otimes \hat{\mathbb{P}}^Y)(y_1, y_2)
\end{aligned}$$

which is the discrete version of the normalising constant α_φ. Therefore we would always obtain an estimate of 1 for $\Lambda_\varphi(Y|X)$.

Essentially this is an overfitting problem: When only one observation of Y is available for each possible value of X then, conditional on the data, knowledge of X does indeed allow perfect prediction of Y, that is, there exists some measurable g, such that $g(x_i) = y_i$ and thus it is not surprising that we obtain a value of 1 for $\Lambda_\varphi(Y|X)$. The classical solution for overfitting is smoothing and for this purpose we will use Checkerboard copulas to obtain an estimator for continuous (X, Y). To justify this approach we need the following

Proposition 4. *Let (X, Y) follow some continuous bivariate distribution H. Let A be the copula of (X, Y) and let K_A be the associated Markov kernel. Then $K(x, (-\infty, y]) := K_A(F_X(x), [0, F_Y(y)])$ is a Markov kernel for (X, Y) and we have*

$$\Lambda_\varphi(Y|X) = \int_0^1 \int_0^1 \int_0^1 \varphi(K_A(u_1, [0, v]) - K_A(u_2, [0, v])) d\lambda(v) d\lambda(u_1) d\lambda(u_2)$$

This means that in the continuous setting $\Lambda_\varphi(Y|X)$ is only dependent on the underlying copula and not on the marginal distributions. Therefore $\Lambda_\varphi(A) := \Lambda_\varphi(Y|X)$ is well defined. Now let $\mathfrak{Cb}_N(A)$ be the N-Checkerboard approximation of the copula A (see e.g. [15,18]). By the *empirical copula* E_n we mean the bilinear interpolation described in [11,12]. Then we have the following

Lemma 1. *Let $(X_1, Y_1), ..., (X_n, Y_n)$ be an i.i.d. sample from some continuous distribution H and let A be the associated copula and let E_n be the empricial copula. Let $N(n) := \lfloor n^s \rfloor$ for some $s \in (0, 1/2)$ then the sequence $(\mathfrak{Cb}_{N(n)}(E_n))_{n\in\mathbb{N}}$ converges weakly conditional to A with probability one as defined in [16].*

Ultimately we arrive at the

Theorem 2. *Let $(X_1, Y_1), ..., (X_n, Y_n)$ be an i.i.d. sample from some continuous distribution H and let A be the associated copula. Let E_n be the empirical copula and let $N(n)$ be defined as in the previous Lemma. Then almost surely we have*

$$\lim_{n\to\infty} \Lambda_\varphi(\mathfrak{Cb}_{N(n)}(E_n)) = \Lambda_\varphi(A).$$

That is, pluging the checkerboard approximation of the empirical copula into Λ_φ results in a strongly consistent estimator.

Funding. The authors gratefully acknowledge the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP). The second and the third author further acknowledge the support of the Austrian Science Fund (FWF) project P 36155-N *ReDim: Quantifying Dependence via Dimension Reduction.*

References

1. Ansari, J., Fuchs, S.: A simple extension of Azadkia and Chatterjee's rank correlation to a vector of endogenous variables. Available at arXiv:2212.01621 (2022)
2. Ansari, J., Langthaler, P.B., Fuchs, S., Trutschnig, W.: Quantifying and estimating dependence via sensitivity of conditional distributions. Available at arXiv:2308.06168 (2023)
3. Azadkia, M., Chatterjee, S.: A simple measure of conditional dependence. Ann. Stat. **49**(6), 3070–3102 (2021)
4. Bickel, P.: Measures of independence and functional dependence. Available at arxivorg/abs/220613663v1 (2022)
5. Chatterjee, S.: A new coefficient of correlation. J. Amer. Statist. Ass. **116**(536), 2009–2022 (2020)
6. Deb, N., Ghosal, P., Sen, B.: Measuring association on topological spaces using kernels and geometric graphs. Available at http://12884418/abs/201001768 (2020)
7. Dette, H., Siburg, K.F., Stoimenov, P.A.: A copula-based non-parametric measure of regression dependence. Scand. J. Statist. **40**(1), 21–41 (2013)
8. Ding, A., Dy, J., Li, Y., Chang, Y.: A robust-equitable measure for feature ranking and selection. J. Mach. Learn. Res. **18**, 1–46 (2017)
9. Fuchs, S.: Quantifying directed dependence via dimension reduction. J. Multivariate Anal. **201**, Article ID 105266 (2024)
10. Gamboa, F., Gremaud, P., Klein, T., Lagnoux, A.: Global sensitivity analysis: a novel generation of mighty estimators based on rank statistics. Bernoulli **28**(4), 2345–2374 (2022)
11. Genest, C., Nešlehová, J.G., Rémillard, B.: Asymptotic behavior of the empirical multilinear copula process under broad conditions. J. Multivariate Anal. **159**, 82–110 (2017)

12. Griessenberger, F., Junker, R., Trutschnig, W.: On a multivariate copula-based dependence measure and its estimation. Electron. J. Statist. **16**, 2206–2251 (2022)
13. Han, F., Huang, Z.: Azadkia-Chatterjee's correlation coefficient adapts to manifold data. Available at arXiv:2209.11156 (2022)
14. Huang, Z., Deb, N., Sen, B.: Kernel partial correlation coefficient—a measure of conditional dependence. J. Mach. Learn. Res. **23**(216), 1–58 (2022)
15. Junker, R., Griessenberger, F., Trutschnig, W.: Estimating scale-invariant directed dependence of bivariate distributions. Comput. Statist. Data Anal. **153**, Article ID 107058 (2020)
16. Kasper, T., Fuchs, S., Trutschnig, W.: On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation. Bernoulli **27**, 2217–2240 (2021)
17. Kinney, J., Atwal, G.: Equitability, mutual information, and the maximal information coefficient. Proc. Natl. Acad. Sci. U.S.A. **111**, 3354–3359 (2014)
18. Li, X., Mikusiński, P., Taylor, M.D.: Strong approximation of copulas. J. Math. Anal. Appl. **225**(2), 608–623 (1998)
19. Nies, T., Staudt, T., Munk, A.: Transport dependency: optimal transport based dependency measures. Available at arXiv:2105.02073 (2023)
20. Shih, J.H., Emura, T.: On the copula correlation ratio and its generalization. J. Multivariate Anal. **182**, Article ID 104708 (2021)
21. Strothmann, C., Dette, H., Siburg, K.: Rearranged dependence measures. Bernoulli **30**(2), 1055–1078 (2024)
22. Sungur, E.A.: A note on directional dependence in regression setting. Comm. Statist. Theory Methods **34**, 1957–1965 (2005)
23. Trutschnig, W.: On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. **384**(2), 690–705 (2011)
24. Wiesel, J.: Measuring association with Wasserstein distances. Bernoulli **28**, 2816–2832 (2022)

Smoke Emission Detection by Deep Learning

Jo-Huan Lee and Chi-Bin Cheng(✉)

Tamkang University, Yingzhuan Rd., Tamsui Dist., New Taipei City 251301, Taiwan
710630079@gms.tku.edu.tw, cbcheng@mail.tku.edu.tw

Abstract. Air pollution has long been a pressing public concern. Before the advent of fixed monitoring equipment, authorities relied solely on human monitoring and citizen reports to collect evidence of pollution events. In an effort to address this issue, the cities of Taichung and Kaohsiung in Taiwan implemented cloud-based smart surveillance systems for automatically detecting and recognizing smoke emission incidents. However, these systems were based on traditional image processing techniques, which limited their detection accuracy. To improve performance, this study introduced a YOLOv7 model into the existing surveillance systems. The YOLO series, based on deep learning techniques, have gained widespread adoption for real-time video analysis tasks due to their high accuracy and efficiency. Moreover, the transfer learning capability of YOLO models allows for fine-tuning on new training samples, enabling adaptation to specific application domains. In this study, the YOLOv7 model was trained on 60,076 labeled photos and subsequently tested on real-world videos collected from the monitoring sites in Taichung and Kaohsiung. The results demonstrated a significant improvement in performance compared to the traditional image processing techniques. Specifically, the overall accuracy of smoke detection increased from 37 to 73%, while the average F1 score improved from 0.33 to 0.64.

Keywords: Smoke emission detection · Image processing · Object recognition · Deep learning · YOLO

1 Introduction

Air pollution issues have long been a topic of public concern. According to statistics by the Ministry of Environment of Taiwan [6], the office received around 279,000 public nuisance and pollution complaint cases every year. The types of cases include air pollution, noise, water pollution, environmental hygiene and waste; among which, air pollution cases accounted for the highest proportion, from 2017 to 2021, with an average of nearly 100,000 cases received per year, accounting for about 36.12% of the total cases. Therefore, the Ministry and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 241–248, 2024.
https://doi.org/10.1007/978-3-031-65993-5_29

local environmental protection bureaus have made efforts to promote relevant policies and work related to air pollution control and air quality management, including driving emission reductions through control strategies for stationary sources, mobile sources, and fugitive sources, and obtaining real-time air quality information through monitoring data.

When air pollution incidents occur, it is often difficult to find the suspect polluters or prove the case due to the difficulty in investigation or evidence collection. This can lead to a vicious cycle of pollution problems. For public perception, such air pollution cases are often distinguished by whether the pollution is visually observable. With the advancement of artificial intelligence in recent years, video recognition applications have made great progress. With surveillance equipment and video transmission via networks, and further video recognition by AI algorithms, automatic recognition of concerned subjects can be achieved. Currently, two cities in Taiwan, Taichung City and Kaohsiung City, have set up surveillance cameras at frequent smoke emission areas, and utilized video recognition algorithms for detecting plums. Video recognition methods are able to assist law enforcement, overcoming problems like insufficient manpower, time-consuming evidence collection, and inability to monitor around the clock.

The recognition tools used in Taichung City and Kaohsiung City are based on computer vision, using dynamic, color and texture algorithms (called Object Contrast Texture Detection, OCTD) to identify smoke plumes. Although the automatic recognition is already capable of focusing on abnormal events, there are still many misjudgments and missed recognitions. The OCTD algorithm has two main drawbacks; firstly, it is too sensitive to produce false alarms often, such as clouds passing being mistaken as smoke plumes; and secondly, the OCTD algorithm often focused on the region of interest (ROI), and does not recognize the smoke at other areas of the image.

In recent years, with the rapid development of deep learning, object detection technologies led by the YOLO series [8] based on deep learning techniques have been widely applied in dynamic or real-time video analysis. After the introduction of YOLOv5 [2], it has again pushed this series to new heights. After commercialization by companies, YOLOv5 features high affinity, efficient algorithms, lightweight architecture and high accuracy. This study attempts to improve the accuracy of smoke emission detection by introducing the YOLO model into the current detection system. Performance of the proposed approach is evaluated by the real cases collected by the surveillance cameras of the Taichung City and Kaohsiung City.

2 Related Work

Smoke detection techniques were mostly used for forest wildfire surveillance traditionally. The advancements in image processing techniques, gradually decreased costs and improved the accuracy of automatic fire alarm detection at the same time. The method is generally to use computer vision programs to convert the RGB color space of images collected by surveillance cameras into

hue, saturation, and intensity, in order to clearly determine the location of smoke plumes [7].

To be able to issue fire alarms earlier, smoke plumes from fires are identified before flames. Smoke plumes are substances generated from combustion reactions, containing mixtures of gases, liquids and solids with air. Smoke plumes are mainly composed of fine carbon particles from complete combustion, and ash from incomplete combustion. Thus, one way to detect smoke plumes is to identify such constituent characteristics. When the detection task is based on static images, color-based decision rules are often used, such as gray scale colors exhibited by the appearance of smoke; while the data source is a dynamic video, continuous and persistent changes in the source of smoke plumes, and changes in volume over time when smoke dissipates, can be used to determine if it is a smoke plume [11].

For fire and smoke detection, the characteristics of fire and smoke had been utilized separately to establish two image computation methods for separate detection of fires. The RGB image characteristics for detecting fire sources and smoke can be used, such as the reddish characteristic of burning fire sources and the light gray characteristic of smoke images, to establish modules based on these image characteristics for fire detection. Results have shown up to 99% correct fire detection rates of such methods [1].

Yang et al. [9] proposed a method for smoke characteristics and segmentation of smoke regions. First, the color distribution regularities of smoke images are enhanced to make the smoke regions stand out more than other regions. Then the Horn-Schunck algorithm is used to calculate the motion energy map of the image. In experiments, this method achieved 93.0% accuracy for smoke segmentation, and up to 99.0% accuracy for forest fire detection.

In addition to basic RGB image characteristics, some studies also attempted to use the characteristics or shapes of smoke for image recognition. Shape conditions can also be added to remove excessively thin smoke shapes and judge appropriate size ranges as smoke regions, otherwise not considered smoke. By setting relevant conditions based on smoke characteristics, wildfire smoke regions can be accurately recognized [5].

In recent years, artificial intelligence or machine learning has begun to be used for smoke detection. For example, to detect smoke emissions by commercial ships anchored at the Keelung Port of Taiwan, Li [4] used machine learning algorithms and data analysis methods to compare images before and after the emissions, with accuracy of 72% to detect the emissions. For the same purpose, Yang [10] used ENVI remote sensing image analysis software to read images with 6 bands (R, G, B, temperature, temperature gradient and red gradient). His approach first obtains multi-spectral image data, then uses a supervised analysis (Minimum Distance) and an unsupervised learning (K-mean clustering), as well as a region of interest (ROI) analysis to build detection models. This method can identify black smoke and filter out lighter gray smoke plumes. Among 34 analyzed images, the method was able to perfectly locate the smoke areas in all black smoke images; while for light smoke images, 3 were misjudged as black smoke, with an accuracy of 86.36%.

Around 55% of the reported air pollution cases each year in Taiwan, relate to visible pollution cases such as smoke from factory chimneys, open area burning, and dust. Pollutant emissions from fixed sources like factory chimneys have long been a focus of public concern. In Taiwan, many cities have set up CCTV near industrial zones or major pollution sources to collect evidences for ascertaining the timing and conditions when air pollution incidents occur. In 2020, Taichung City Environmental Protection Bureau implemented a cloud-based smart surveillance system at air pollution hotspots. These areas often received public complaints about air pollution or abnormal emissions from factories. In 2021, Kaohsiung City Environmental Protection Bureau also introduced the same system and set up cameras at its industrial parks. The installed cloud-based video surveillance and recognition systems are able to continuously identify abnormalities from real-time images, and automatically recognize events like chimney emissions, fires, factory area abnormal emissions, open burning, and dust raising in unattended.

Since Sep 2020 when Taichung City's system was activated, a total of 41 air pollution events have been discovered through the surveillance system, with 10 cases filed and penalized after inspection; the fines in total was around 200 thousand USD. Kaohsiung City's system discovered 47 air pollution events, with 8 cases penalized after inspection; total fines so far exceed 70 thousand USD [3]. However, as mentioned earlier in the previous section, the cloud-based smart surveillance system still suffers from many missed recognitions.

3 Proposed Methods

This study attempts to improve the accuracy of smoke emission detection by introducing the YOLO model into the current cloud-based smart surveillance system which is based on OCTD algorithm. The following sub-sections first introduces the OCTD algorithm, and then presents the YOLO model used for the same task.

3.1 OCTD Algorithm

The procedures of the OCTD algorithm are as follows. The received video stream is first decoded, and then the image frame is analyzed through three modules, moving object detection, color characteristics detection, and texture detection, and the results reported in both text and images with labeled objects.

When identifying motion in an image, after calculating the motion vector, the similarity between this vector and the various trained motion models is calculated. The motion model with the highest similarity is taken as the recognition result. The color characteristic detection model is based on the idea that smoke is usually presented in a single color. The original RGB space is converted to HSV space and a single color object is assessed when there is a large area and persistently existing object in the image. Texture refers to the probability distribution and geometric arrangement of gray-level values of objects in an image, or

patterns formed by certain regularities. The idea is the smoke flume is generally with smooth texture. The thresholds for the decisions of the above three modules are determined by testing and adjusting manually from real case images to assess their effectiveness. The determined thresholds are set as parameters to the three modules and continuously checked and adjusted according to their performances.

3.2 YOLO Model

YOLO (You Only Look Once) is an object recognition neural network model that can simultaneously detect multiple object locations and recognize the objects by inputting the entire image into a convolutional neural network (CNN), with the advantage of enabling real-time detection while maintaining high accuracy. YOLO divides the input image into many grids, then detects if the center point of each cell falls into a certain grid, in which case that grid is responsible for detecting the target object; in each grid, bounding boxes and confidence scores are predicted.

YOLO object recognition is one-staged, performing object classification and localization in the same neural network. The image input is passed through this neural network for recognition and output, predicting bounding box coordinates, object confidence scores, and object class probabilities. It has now been developed up to YOLOv8 V (released on 2023/1/10). The version used in this study is YOLOv7, which was released in July 22 of 2022, reducing model parameters by about 40% and computational operations by 50% compared to its previous versions.

The YOLO series own transfer learning capabilities; that is, they are able to transfer previously trained object feature information to a new model for more effective and accelerated learning, rather than training the model from scratch. Transfer learning allows users applying YOLO to their own problem domains by fine-tuning the model with newly added training samples.

To fulfill our goal of smoke emission detection, we manually labeled the images collected from the cloud-based smart surveillance systems with two classes, black smoke and white smoke, which are the demands of the Environmental Protection Bureaus of the two cities. When none of the two classes is detected from the image, it is assessed as the third class: none smoke.

A total of 60,076 labeled photos were used (15,123 black smoke, 33,188 white smoke, and 11,765 none smoke). Of which, 18% of the dataset were reserved as the test set, and the rest as the training and validation sets. The hardware used was CPU of i9-12900 16-core/24-thread, GPU of Nvidia RTX 3090 24G, with 64G RAM. The number of training epochs was set to 300 based on our experience that at least 300 epochs were needed for the algorithm's performance to meet the goal. The training time was around 150 h.

To confirm the model's accuracy after training, the reserved test set photos were used, containing 2043 black smoke, 4667 white smoke and 4103 none smoke photos. The precision rates are 0.60 and 0.62 for black and white smokes respectively, and the overall F1 score for all classes is 0.67.

YOLO uses a confidence measure, which is the model's confidence that a bounding box contains a particular object, to estimate the prediction accuracy. In our application, the confidence threshold is set as 0.35, which is determined empirically through multiple training runs, to determine which class of the smoke exists in the photo. For online application, the sampling frame rate is 15 FPS (Frames per second), and the recognition task is performed once every 2 frames. If 20 out of 90 frames have the same recognition result, an alert event is triggered.

4 Performance Evaluation

We randomly selected 230 event videos from monitoring sites in Taichung and Kaohsiung cities between July and September 2023, in which, 189 videos were taken in day times and 41 were taken at night. The descriptions of the above videos are presented in Table 1. Among the 230 videos, there were 35 black smoke events, accounting for 15% of all events, 168 white smoke events, accounting for 73%, and 27 none smoke events, accounting for 12%. This distribution is close to the total cases found by the cloud-based smart surveillance system; generally, fewer black smoke events while much more white smoke events.

Table 1. Descriptions of the collected event videos.

				Visibility index			
Period	Video amount	Number of true cases		Avg.	Std. dev.	Max	Min
Day	189	White	128	906.42	412.74	272	2017
		Black	35	815.68	683.99	166	2618
		None	26	936.46	534.15	120	1942
Night	141	White	40	1034.05	327.34	248	1523
		Black	0	-	-	-	-
		None	1	528.00	-	-	-

The visibility index in Table 1 serves as an initial screening of the videos. This index indicates the clarity degree of the video to be able identified by the system. The computation of the visibility index V is based on the variance of gray-scale values of the image as defined below.

$$V = \frac{\Sigma_i \Sigma_j (p_{ij} - \bar{p_{i.}})^2}{I \cdot J}$$

where p_{ij} is the gray value of the pixel at the position row i and column j of the image, $\bar{p_{i.}}$ is the average gray value of all pixels at row i, and I and J are the numbers of row and column respectively. The average, standard deviation, maximum and minimum visibility indexes of videos are also shown in Table 1. In our experience, images with a visibility index lower than 120 can barely be identified by the system, and thus are excluded from the detection process.

4.1 Performance of OCTD

The confusion matrix of the detection by OCTD is shown in Table 2. From Table 2 we could compute the overall accuracy is 37%, the precisions for the three classes (black, white and none) are 6%, 73% and 0% respectively, the recalls are 17%, 46% and 0% respectively, and F1 scores are 9 and 57% for black smoke and white smoke (not applicable to none smoke due to 0 precision). From Table 2 we find that there are many cases of misjudging white smoke as black smoke. It is likely attributed to the binarized images in the process losing the color information.

Table 2. Confusion matrix of the detection by OCTD

		Predicted			
		Black	White	None	
Actual	Black	6	9	20	35
	White	90	78	0	168
	None	7	20	0	27
		103	107	20	230

4.2 Performance of YOLO

The confusion matrix of the detection by YOLO is shown in Table 3. From Table 3 we could compute the overall accuracy is 73%, the precisions for the three classes (black, white and none) are 100%, 87% and 37% respectively, the recalls are 40%, 77% and 93% respectively, and F1 scores are 57%, 81% and 53% respectively. Comparing the results by the two methods, YOLO has greatly improved the detection performance by OCTD.

Table 3. Confusion matrix of the detection by YOLO

		Predicted			
		Black	White	None	
Actual	Black	14	18	3	35
	White	0	129	39	168
	None	0	2	25	27
		14	149	67	230

5 Conclusion

This study introduces artificial intelligence into the cloud-based smart surveillance system currently used by Taichung and Kaohsiung cities of Taiwan for detecting smoke emissions. By adopting and fine-tuning the model with new training samples, we are able to improve the performance of the traditionally used OCTD algorithm, which is generally based on image processing techniques. To evaluate the performance of the new method, 230 event videos from monitoring sites in the two cities between July and September 2023 were collected to test the two methods. The overall accuracy of detection is improved from 37% to 73%, and F1 score is improved from 0.33 to 0.64 on average.

Though YOLO outperforms OCTD greatly, it is noted that many cases of white smoke were recognized as none smoke by YOLO, while it was never happened by OCTD. In our future study, we attempt to integrate the two methods with the expectation to remedy such cases.

References

1. Çelik, T., Özkaramanlı, H., Demirel, H.: Fire and smoke detection without sensors: image processing based approach. In: Signal Processing Conference 15th European, IEEE, pp. 1794–1798 (2007)
2. Jocher, G.: YOLOv5 (2022). Available at https://github.com/ultralytics/yolov5
3. Lee, J.-H.: Application of Artificial Intelligence Object Recognition Techniques in Smoke Emission Pollution Detection. Master's Thesis, Department of Information Management, Tamkang University (2024)
4. Li, S., Wang, B., Dong, R., Zhou, Z., Li, S.: A novel smoke detection algorithm based on fast self-tuning background subtraction. In: 2016 Chinese Control and Decision Conference (CCDC), Yinchuan, pp. 3539–3543 (2016). https://doi.org/10.1109/CCDC.2016.7531596
5. Li, X.-R.: Detecting ship smoke from the images of the Keelung harbor. Master's Thesis, Department of Marine Environment and Information Science, National Taiwan Ocean University (2016)
6. Ministry of Environment of Taiwan: Number of Public Nuisance Complaint Cases (2024). Available at https://claude.ai/chat/19f5120c-d0c7-46e5-bb49-dad24f200e6b
7. Pu, Y.-R., Wang, P.-C.: A feasibility study on smoke identification in the workplace—dynamic analysis using color saturation. Ind. Safety Health Monthly **259**, 41–48 (2011)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
9. Yang, J., Yuan, J., Wang, J., Fang, J., Zhang, Y.: A saliency-based method for early smoke detection in video sequences. Fire Technol. **52**(5), 1271–1292 (2016)
10. Yang, T.-J.: Smoke Detection and Analysis of Ship Images in Keelung Harbor. Master's Thesis, Department of Marine Environment and Information Science, National Taiwan Ocean University (2018)
11. Yin, Y.-H.: Video-based Early Fire Smoke Detection. Master's Thesis, Graduate Institute of Electronic and Information Engineering, National Kaohsiung University of Applied Sciences (2007)

Quantifying Directed Dependence with Kendall's Tau

Carsten Limbach(✉) and Sebastian Fuchs

Department for Artificial Intelligence and Human Interfaces, University of Salzburg, Salzburg, Austria
{carsten.limbach,sebastian.fuchs}@plus.ac.at

Abstract. The Markov product of copulas is a powerful tool for modeling directed dependence between two random variables X and Y, and its evaluation via certain concordance measures such as Spearman's rho and Spearman's footrule allows for quantifying different aspects of directed dependence, and thus, different types of influence that X can exert on Y. Motivated by the recent publications [3,11], we apply Kendall's tau to the Markov product which leads to a dependence measure determining the degree of agreement among the conditional distributions of Y given $X = x$.

Keywords: Copulas · Directed dependence · Markov product · Kendall's tau

1 Introduction

Directed dependence describes the influence or effect that one variable has on another variable and/or vice versa, and the knowledge of such an interaction is essential for understanding complex data structures. Unlike undirected dependence evaluated for instance via concordance measures such as Spearman's rho and Kendall's tau and which merely indicates the existence of a relationship without clarifying its direction, directed dependence instead allows for investigating causality, thereby providing insights into the extent to which changes in variable X can lead to or contribute to changes in variable Y. In a regression type setting, X is denoted as *predictor variable* and Y is called *response variable*.

In this paper, directed dependence is explored through the lens of the well-known Markov product of copulas; see, e.g., [1,2]. The basic idea underlying this mapping is to extend the random vector (X, Y) to (X, Y, Y') with Y' being an independent copy of Y sharing the same conditional distribution and being conditionally independent given X, and then to restrict oneself exclusively to evaluating the dependence between Y and Y'. For an illustration, let us consider the two extreme cases: If Y strongly depends on X then, for two data points with only slightly differing X values, the respective Y values should also differ only slightly. If, instead, X and Y are independent, then the concrete

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 249–255, 2024.
https://doi.org/10.1007/978-3-031-65993-5_30

choice of the X values has no effect on the corresponding Y values.

The random variable Y is said to be *perfectly dependent* on the random variable X, if there exists a measurable function h such that $Y = h(X)$ almost surely; cf., e.g., [3].

Evaluating the Markov product and hence the vector (Y, Y') through the various established dependence measures opens up numerous approaches for quantifying directed dependence, each of which allows for detecting a different aspect of directed dependence; we refer to [3,11] in which the Markov product has been evaluated via Spearman's footrule, Spearman's rho and Gini's gamma. Motivated by the theoretical results in [11], in this work we employ Kendall's tau to the Markov product and present explicit formulas for the case the underlying dependence structure belongs to either a Gaussian copula, a Marshall-Olkin copula or a Fréchet copula.

We will make use of the following definition [4,10]: Given a pair of random variables (Z_1, Z_2) with continuous marginal distribution functions and connecting copula C, Kendall's tau is defined as the difference between the probability of concordance and the probability of discordance, i.e.,

$$\tau(Z_1, Z_2) := \mathbb{P}\left((Z_1 - Z_1^*)(Z_2 - Z_2^*) > 0\right) - \mathbb{P}\left((Z_1 - Z_1^*)(Z_2 - Z_2^*) < 0\right), \quad (1)$$

where (Z_1^*, Z_2^*) denotes an independent copy of (Z_1, Z_2). Kendall's tau viewed as a mapping on copulas $\mathcal{C} \to \mathbb{R}$ then fulfils

$$\tau(C) = 4 \int_{[0,1]^2} C(u, v)\, \mathrm{d}\mu_C(u, v) - 1\,, \quad (2)$$

where $\mathcal{C}$ denotes the collection of all (bivariate) copulas and μ_C denotes the probability measure corresponding to the copula C. Kendall's tau equals 1 if and only if Z_1 and Z_2 are comonotonic, i.e., the corresponding copula C coincides with the upper Fréchet-Hoeffding bound M given by $M(s, t) = \min\{s, t\}$. In addition, Kendall's tau equals 0 when Z_1 and Z_2 are independent, i.e., the corresponding copula C coincides with the independence copula Π given by $\Pi(u, v) = uv$.

2 A Copula for Investigating Directed Dependence

Consider the pair of random variables (X, Y) with continuous marginal distribution functions F of X and G of Y and connecting copula A. Its extension (X, Y, Y') with Y' being an independent copy of Y sharing the same conditional distribution and being conditionally independent given X then fulfils

$$\begin{aligned}\mathbb{P}(Y \le y, Y' \le y') &= \int_{\mathbb{R}} \mathbb{P}(Y \le y, Y' \le y' \,|\, X = x)\, \mathrm{d}\mathbb{P}^X(x) \\ &= \int_{\mathbb{R}} \mathbb{P}(Y \le y \,|\, X = x)\, \mathbb{P}(Y' \le y' \,|\, X = x)\, \mathrm{d}\mathbb{P}^X(x)\,.\end{aligned}$$

The copula of (Y, Y') coincides with the Markov product of A being defined as

the mapping $\psi : \mathcal{C} \to \mathcal{C}$ given by

$$\psi(C)(s,t) := \int_{[0,1]} K_C(u,[0,s])\, K_C(u,[0,t])\, \mathrm{d}\lambda(u)\,, \tag{3}$$

where K_C denotes the Markov kernel[1] of copula C. The next result justifies the use of the Markov product for investigating directed dependence.

Proposition 1. *[3, Thm. 1] Consider the random vector (X,Y) with continuous marginals and connecting copula A.*

1. *The following statements are equivalent:*
   (a) Y is perfectly dependent on X.
   (b) $\psi(A) = M$.
2. *The following statements are equivalent:*
   (a) X and Y are independent.
   (b) $\psi(A) = \Pi$.

We now illustrate the impact of the transformation ψ by applying it to some well-known parametric copula families.

Example 1. *[3] (Gaussian copulas)*
Consider the Gaussian copula A_ρ with (correlation) parameter $\rho \in (-1,1)$ and recall that $A_0 = \Pi$. Then $\psi(A_\rho)$ is again Gaussian with parameter $\rho^2 < |\rho|$, i.e., $\psi(A_\rho) = A_{\rho^2}$, implying that the class of Gaussian copulas is closed under ψ. Figure 1 depicts a sample of size $n = 10,000$ taken from a Gaussian copula A_ρ with parameter $\rho = -\sqrt{0.5}$ along with a sample of equal size from $\psi(A_\rho)$: In addition to a shift towards the identity line, a broader spread can be observed. This so-called smoothing effect for the Markov product applied to stochastically increasing[2] copulas is well recognised; see, e.g., [12].

Example 2. *[3] (Marshall-Olkin copulas)*
For $\alpha, \beta \in [0,1]$, consider the Marshall-Olkin copula $A_{\alpha,\beta}$ given by

$$A_{\alpha,\beta}(u,v) := \min\{u^{1-\alpha}v, uv^{1-\beta}\}\,.$$

[1] A Markov kernel from $[0,1]$ to $\mathcal{B}([0,1])$ is a mapping $K : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ such that for every fixed $F \in \mathcal{B}([0,1])$ the mapping $u \mapsto K(u,F)$ is measurable and for every fixed $u \in [0,1]$ the mapping $F \mapsto K(u,F)$ is a probability measure. Given uniformly distributed random variables U and V on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we say that a Markov kernel K is a regular conditional distribution of V given U if $K(U(\omega), F) = \mathbb{P}(V \in F \,|\, U)(\omega)$ holds $\mathbb{P}$-almost surely for every $F \in \mathcal{B}([0,1])$. It is well-known that for each such random vector (U,V) a regular conditional distribution $K(.,.)$ of V given U always exists and is unique for $\mathbb{P}^U$-a.e. $u \in [0,1]$, where $\mathbb{P}^U$ denotes the push-forward of $\mathbb{P}$ under U. For more background on conditional expectation and general disintegration we refer to [6,8]; for more information on Markov kernels in the context of copulas we refer to [2,7,9].

[2] A copula C is said to be *stochastically increasing* if, for (a version of) the Markov kernel K_C and any $v \in (0,1)$, the mapping $u \mapsto K_C(u,[0,v])$ is non-increasing; see, e.g., [5].

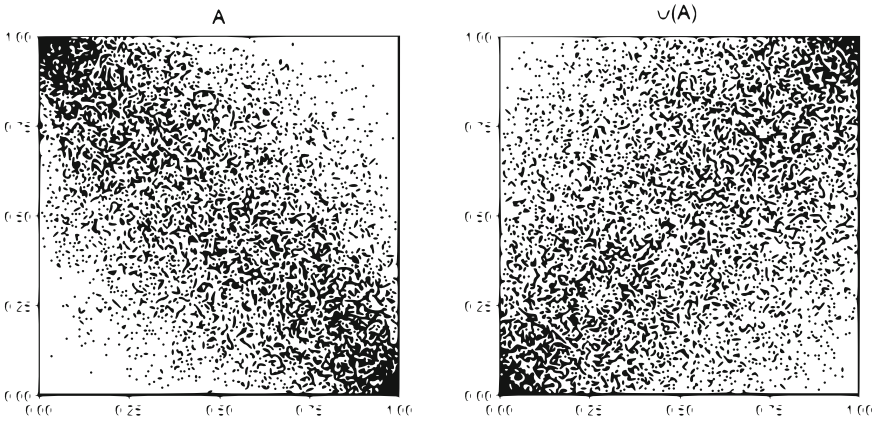


Fig. 1. Samples of size $n = 10,000$ drawn from a Gaussian copula with correlation parameter $\rho = -\sqrt{0.5}$ (left panel) and its Markov product (right panel) that is Gaussian with parameter $\rho^2 = 0.5$.

Recall that $\Pi \leq A_{\alpha,\beta} \leq M$, $A_{\alpha,\beta} = \Pi$ *if and only if* $\min\{\alpha,\beta\} = 0$, *and* $A_{\alpha,\beta} = M$ *if and only if* $\min\{\alpha,\beta\} = 1$. *Its Markov product fulfils*

$$\psi(A_{\alpha,\beta})(s,t) = \begin{cases} \Pi(s,t) + \frac{\alpha^2}{1-2\alpha}\Pi(s,t)\left(1 - \Pi(s,t)^{\frac{\beta-2\alpha\beta}{\alpha}} M(s,t)^{\frac{2\alpha\beta-\beta}{\alpha}}\right) & \alpha \notin \{0, \frac{1}{2}, 1\} \\ \Pi(s,t) & \alpha = 0 \\ \Pi(s,t) + \frac{\beta}{2}\Pi(s,t)\left(\log(M(s,t) - \log(\Pi(s,t))\right) & \alpha = \frac{1}{2} \\ A_{\beta,\beta}(s,t) & \alpha = 1 \end{cases}$$

so that $\psi(A_{\alpha,\beta})$ *is again Marshall-Olkin whenever* $\min\{\alpha,\beta\} = 0$ *or* $\alpha = 1$*; in all other cases, the Markov product of a Marshall-Olkin turns out to be a mixture of its extremes* Π *and* M.

Fig. 2 depicts a sample of size $n = 10,000$ *taken from a Marshall-Olkin copula* $A_{\alpha,\beta}$ *with parameter vector* $(\alpha,\beta) = (0.8, 0.2)$ *along with a sample of equal size from* $\psi(A_{\alpha,\beta})$*: The sample on the left exhibits a clustering of data points at* $v = u^{\beta/\alpha}$, *and the Markov product shifts this singular part of the copula to the identity line. It can also be seen from the sample on the right that the identity line is more pronounced for high* Y *values, indicating that the Markov product can be used not only for visualising the extent of dependence but also for drawing concrete conclusions about the regions in which this dependence is particularly strong.*

Example 3. *[3] (Fréchet copulas)*
For $\alpha,\beta \in [0,1]$ *with* $\alpha + \beta \leq 1$, *consider the Fréchet copula* $A_{\alpha,\beta}$ *given by*

$$A_{\alpha,\beta} := \alpha\, M + (1 - \alpha - \beta)\, \Pi + \beta\, W\,.$$

Clearly, $A_{\alpha,\beta} = \Pi$ *if and only if* $\alpha = 0 = \beta$, $A_{\alpha,\beta} = M$ *if and only if* $\alpha = 1$, *and* $A_{\alpha,\beta} = W$ *if and only if* $\beta = 1$. *Its Markov product fulfils*

$$\psi(A_{\alpha,\beta}) = A_{\alpha^*,\beta^*}\,,$$

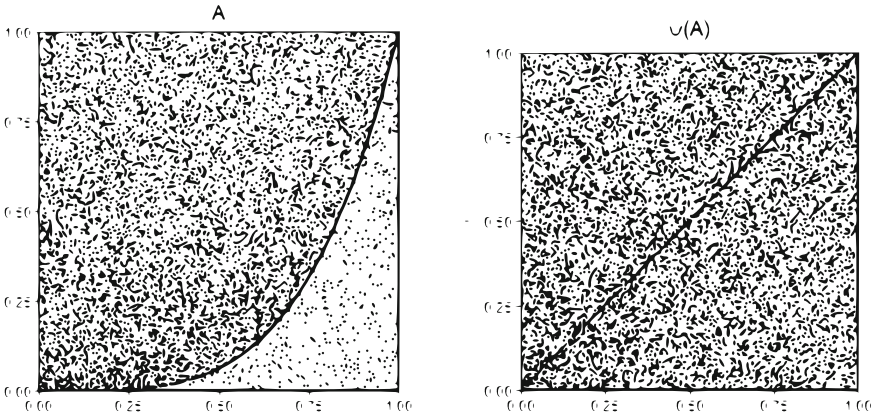


Fig. 2. Samples of size $n = 10,000$ drawn from a Marshall-Olkin copula with parameter vector $(\alpha, \beta) = (0.8, 0.2)$ (left panel) and its Markov product (right panel).

where $\alpha^* = \alpha^2 + \beta^2$ *and* $\beta^* = 2\alpha\beta$*, implying that the class of Fréchet copulas is closed under* ψ *with the three fixed points* Π *(*$\alpha = 0 = \beta$*),* M *(*$\alpha = 1$*) and* $(M + W)/2$ *(*$\alpha = 1/2 = \beta$*). Figure 3 depicts a sample of size* $n = 10,000$ *taken from a Fréchet copula* $A_{\alpha,\beta}$ *with parameter vector* $(\alpha, \beta) = (0.1, 0.1)$ *along with a sample of equal size from* $\psi(A_{\alpha,\beta})$*.*

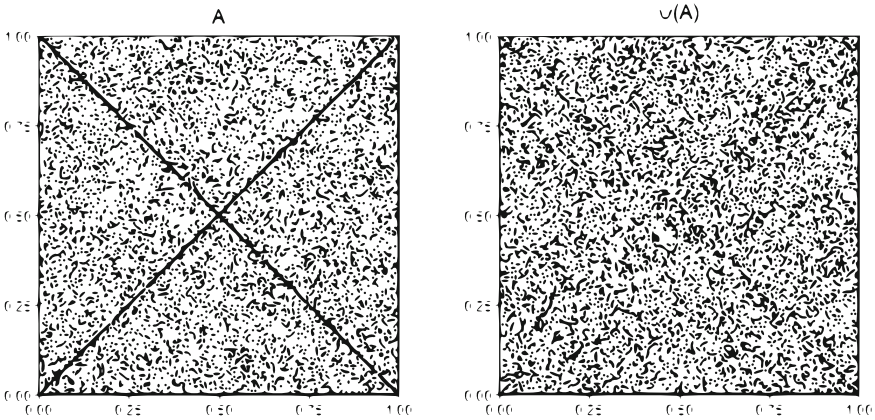


Fig. 3. Samples of size $n = 10,000$ drawn from a Fréchet copula copula with parameter vector $(\alpha, \beta) = (0.1, 0.1)$ (left panel) and its Markov product (right panel).

3 Kendall's Tau and the Markov Product

Combining Kendall's tau with the Markov product allows for determining the degree of agreement among the conditional distributions of Y given $X = x$.

We define the mapping $\mathcal{T}$ for random vectors (X, Y) with connecting copula A by letting

$$\mathcal{T}(Y|X) := \tau(\psi(A)), \tag{4}$$

and highlight the following result which is due to [11].

Proposition 2. *[11, Thm. 3]*
Consider the random vector (X, Y) with continuous marginals and connecting copula A. Then $\mathcal{T}$ fulfils

$$\mathcal{T}(Y|X) = \frac{\mathrm{Var}\left(\mathbb{E}\left(\mathbb{1}_{\{Y<Y^*\}} \mid X, X^*\right)\right)}{\mathrm{Var}\left(\mathbb{1}_{\{Y<Y^*\}}\right)}$$

where (X^*, Y^*) denotes an independent copy of (X, Y).

The next result sheds light on the extreme cases.

Proposition 3. *[11, Thm. 4], [3, Cor. 3]*
Consider the random vector (X, Y) with continuous marginals and connecting copula A. Then

1. $\mathcal{T}(Y|X) = 0$ if and only if $\mathbb{P}(Y < Y^* \mid X, X^*) = \frac{1}{2}$ almost surely.
2. $\mathcal{T}(Y|X) = 1$ if and only if Y is perfectly dependent on X.

where (X^*, Y^*) denotes an independent copy of (X, Y).

We now resume the parametric copula families discussed in Examples 1 to 3 and provide explicit expression for $\mathcal{T}(Y|X)$.

Example 4. *(Gaussian copulas)*
Consider the Gaussian copula A_ρ with (correlation) parameter $\rho \in (-1, 1)$. According to Example 1, $\psi(A_\rho)$ is again Gaussian with parameter $\rho^2 < |\rho|$ and (4) yields

$$\mathcal{T}(Y|X) = \tau(A_{\rho^2}) = \frac{2}{\pi} \arcsin(\rho^2).$$

Thus, $\mathcal{T}(Y|X) = 0$ if and only if $\rho = 0$, and $\lim_{\rho \to 1} \mathcal{T}(Y|X) = 1$.

Example 5. *(Marshall-Olkin copulas, $\alpha = 1$)*
For $\alpha = 1$ and $\beta \in [0, 1]$, consider the Marshall-Olkin copula $A_{\alpha,\beta}$. According to Example 2, $\psi(A_{\alpha,\beta})$ is again Marshall-Olkin with parameter vector (β, β) and (4) yields

$$\mathcal{T}(Y|X) = \tau(A_{\beta,\beta}) = \frac{\beta}{2 - \beta}.$$

Thus, $\mathcal{T}(Y|X) = 0$ if and only if $\beta = 0$, and $\mathcal{T}(Y|X) = 1$ if and only if $\beta = 1$.

Example 6. *(Fréchet copulas)*
For $\alpha, \beta \in [0,1]$ with $\alpha + \beta \leq 1$, consider the Fréchet copula $A_{\alpha,\beta}$. According to Example 3, $\psi(A_{\alpha,\beta})$ is again a Fréchet copula with parameter vector $\alpha^* = \alpha^2 + \beta^2$ and $\beta^* = 2\alpha\beta$, and (4) yields

$$\mathcal{T}(Y|X) = \tau(A_{\alpha^2+\beta^2,2\alpha\beta}) = \frac{1}{3}(\alpha^2 - \beta^2)^2 + \frac{2}{3}(\alpha - \beta)^2 .$$

Thus, $\mathcal{T}(Y|X) = 0$ if and only if $\alpha = \beta$, and $\mathcal{T}(Y|X) = 1$ if and only if $(\alpha, \beta) \in \{\{0,1\},\{1,0\}\}$.

Acknowledgements. Both authors gratefully acknowledge the support of the Austrian Science Fund (FWF) project P 36155-N *ReDim: Quantifying Dependence via Dimension Reduction* and the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

References

1. Darsow, W., Nguyen, B., Olsen, E.: Copulas and Markov processes. Illinois J. Math. **36**, 600–642 (1992)
2. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton FL (2016)
3. Fuchs, S.: Quantifying directed dependence via dimension reduction. J. Multivariate Anal. **201**, 105266 (2024)
4. Fuchs, S., Schmidt, K.D.: On order statistics and Kendall's tau. Stat. Probab. Lett. **169**, 108972 (2021)
5. Fuchs, S., Tschimpke, M.: Total positivity of copulas from a Markov kernel perspective. J. Math. Anal. Appl. **518**, 126629 (2023)
6. Kallenberg, O.: Foundations of Modern Probability. Springer, New York (2002)
7. Kasper, T., Fuchs, S., Trutschnig, W.: On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation. Bernoulli **27**, 2217–2240 (2021)
8. Klenke, A.: Wahrscheinlichkeitstheorie. Springer, Heidelberg (2008)
9. Mroz, T., Fuchs, S., Trutschnig, W.: How simplifying and flexible is the simplifying assumption in pair-copula constructions - analytic answers in dimension three and a glimpse beyond. Electron. J. Stat. **15**, 1951–1992 (2021)
10. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)
11. Shih, J., Emura, T.: On the copula correlation ratio and its generalization. J. Multivariate Anal. **182**, 104708 (2021)
12. Siburg, K.F., Strothmann, C.: Stochastic monotonicity and the Markov product for copulas. J. Math. Anal. Appl. **503**, 125348 (2021)

Some Notes on Bivariate Lower Semilinear Copulas

Lea Maislinger(✉) and Wolfgang Trutschnig

Department for Artificial Intelligence and Human Interfaces, University of Salzburg, Salzburg, Austria
lea.maislinger@plus.ac.at, wolfgang@trutschnig.net

Abstract. We revisit the class of bivariate lower semilinear copulas (LSL) as introduced by Durante et al. in 2008 (see [2]), view LSL copulas from the Markov kernel perspective, and use the kernel to derive a simple formula for the singular mass of arbitrary LSL copulas. Complementing these results we derive simple formulas for Kendall's τ and Spearman's ρ and study monotonicity/ordering properties like positive quadrant dependence (PQD), left tail dependence (LTD) and stochastic increasingness (SI).

Keywords: Semilinear copulas · Markov kernel · Singular mass · Concordance · Monotonicity

1 Introduction

To the best of our knowledge the family of (lower and upper) bivariate semilinear copulas was first studied in 2008 by Durante et al. in [2]. In a more recent paper (see [4]) Durante et al. studied extreme points of the compact and convex family of bivariate lower semilinear copulas (LSL copulas, for short) and provided a characterization of extreme points in terms of properties of the corresponding diagonals.
Multivariate versions of upper semilinear copulas were studied by Arias-García et al. in 2016 (see [1]) and by Sloot et al. in 2020 (see [10]), where it was also shown that multivariate upper semilinear copulas are quite pathological in the sense that they concentrate their mass on finitely many hyperplanes (and hence are singular).

Motivated by their simple analytic form and the simplicity of their construction via given copula diagonals δ with specific properties, we here revisit bivariate LSL copulas and provide some small, seemingly novel results: We derive a simple expression for the Markov kernel of LSL copulas and use this expression to show that the singular mass $sing(S_\delta)$ of every LSL copula S_δ is given by

$$sing(S_\delta) = 2 \int_{[0,1]} \frac{\delta(x)}{x} \, d\lambda(x) - 1. \tag{1}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 256–263, 2024.
https://doi.org/10.1007/978-3-031-65993-5_31

Moreover simple expressions for Kendall's τ and Spearman's ρ of S_δ are stated and some standard monotonicity/ordering properties like positive quadrant dependence (PQD), left tail dependence (LTD) and stochastic increasingnes (SI) are studied.

2 Notation and Preliminaries

For every metric space (S, d) the Borel σ-field on S will be denoted by $\mathcal{B}(S)$. In the sequel $\mathcal{C}$ denotes the family of all bivariate copulas. For every $C \in \mathcal{C}$ the corresponding doubly stochastic measure μ_C on $\mathcal{B}([0,1]^2)$ is defined by $\mu_C([0,x] \times [0,y]) = C(x,y)$ for all $x, y \in [0,1]$ (and extended to full $\mathcal{B}([0,1]^2)$ in the standard way). A copula C is called lower semilinear (LSL) if, for all $x \in (0,1]$, the mappings $t \mapsto h_x(t) := C(t,x)$ and $t \mapsto v_x(t) := C(x,t)$ are linear. It is well known that a mapping $\delta : [0,1] \to [0,1]$ is the diagonal section of a copula if, and only if the following four properties hold:

- $\delta(0) = 0$, $\delta(1) = 1$;
- $\delta(u) \leq u$ for all $u \in [0,1]$;
- δ is non-decreasing and
- δ is 2-Lipschitz, i.e., $|\delta(v) - \delta(u)| \leq 2|v - u|$ holds for all $u, v \in [0,1]$.

The set of all diagonals (i.e., of all functions $\delta : [0,1] \to [0,1]$ fulfilling the properties above) will be denoted by $\mathcal{D}$. For more background on copulas we refer to [3,7].
The Lebesgue decomposition of a doubly stochastic measure μ_C with respect to the two-dimensional Lebesgue measure λ_2 will be denoted by

$$\mu_C = \mu_C^{\ll} + \mu_C^{\perp},$$

where $\mu_C^{\ll}$ is the absolutely continuous and $\mu_C^{\perp}$ the singular component of μ_C. The total singular mass $\mu_C^{\perp}([0,1]^2) \in [0,1]$ of $C \in \mathcal{C}$ will be denoted by $sing(C)$, i.e., $sing(C) := \mu_C^{\perp}([0,1]^2) \in [0,1]$.
A mapping $K : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ will be called a Markov kernel if $x \mapsto K(x,B)$ is measurable for every fixed $B \in \mathcal{B}([0,1])$ and $B \mapsto K(x,B)$ is a probability measure on $\mathcal{B}([0,1])$ for every fixed $x \in [0,1]$. It is well known that for every copula $C \in \mathcal{C}$ there exists a Markov kernel $K_C : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ fulfilling

$$\int_E K_C(x,F) d\lambda(x) = \mu_C(E \times F)$$

for all $E, F \in \mathcal{B}([0,1])$. This Markov kernel is unique for λ-almost every $x \in [0,1]$. Vice versa, every Markov kernel $K : [0,1] \times \mathcal{B}([0,1]) \to [0,1]$ fulfilling the fixed-point property

$$\int_{[0,1]} K_C(x,F) d\lambda(x) = \lambda(F)$$

for every $F \in \mathcal{B}([0,1])$ can be shown to be the Markov kernel of a unique copula. For more background on Markov kernels and disintegration we refer to [5], for more results on the interplay between Markov kernels and copulas, e.g., to [6] and [11].

3 Markov Kernels of Bivariate LSL Copulas

We start with the following handy characterization of lower-semilinear copulas (directly following from linearity of the horizontal and vertical sections) going back to [2]:

Theorem 1 ([2]). *For a diagonal $\delta \in \mathcal{D}$, let the function $S_\delta : [0,1]^2 \to [0,1]$ be defined by*

$$S_\delta(x,y) := \begin{cases} y\frac{\delta(x)}{x} & \text{if } y \leq x, \\ x\frac{\delta(y)}{y} & \text{otherwise,} \end{cases} \tag{2}$$

with the convention $\frac{0}{0} := 0$. Then S_δ is a lower semilinear copula if, and only if, the functions $\varphi_\delta, \eta_\delta : (0,1] \to [0,1]$, defined by

$$\varphi_\delta(x) := \frac{\delta(x)}{x}, \eta_\delta(x) := \frac{\delta(x)}{x^2}$$

are non-decreasing and non-increasing, respectively.

In order to underline symmetry of LSL copulas in [4] Durante et al. expressed Eq. (2) as

$$S_\delta(x,y) = \frac{(x \wedge y)\delta(x \vee y)}{x \vee y},$$

where $x \wedge y = \min\{x,y\}$ and $x \vee y = \max\{x,y\}$.
In the sequel we will denote the family of all diagonals of lower semilinear copulas by $\mathcal{D}^{LSL}$, i.e.

$$\mathcal{D}^{LSL} := \left\{ \delta \in \mathcal{D} : x \mapsto \frac{\delta(x)}{x} \text{ non-decreasing}, x \mapsto \frac{\delta(x)}{x^2} \text{non-increasing} \right\},$$

and write $\mathcal{C}^{LSL}$ for the class of all LSL copulas. Lipschitz continuity of diagonals implies λ-almost everywhere existence of the derivative δ'. According to [2] diagonals of LSL copulas fulfill the following useful inequality, which we will use for proving Theorem. 2.

Corollary 1 ([2]). *Suppose that $\delta \in \mathcal{D}$. Then the function S_δ defined according to Eq. (2) is a LSL copula if, and only if*

$$\frac{\delta(x)}{x} \leq \delta'(x) \leq 2\frac{\delta(x)}{x}, \tag{3}$$

holds for all points $x \in (0,1)$ where $\delta'(x)$ exists.

Given some $\delta \in \mathcal{D}^{LSL}$, in the sequel we will let $\hat{w}_\delta : [0,1] \to [0,2]$ denote a Borel-measurable function fulfilling $\delta'(x) = \hat{w}_\delta(x)$ for λ-almost every $x \in [0,1]$ (for a justification of the existence of such a function see, e.g., [8]). Providing a simple expression for the Markov kernel of a LSL copula S_δ we will work with a special version w_δ of $\hat{w}_\delta$ assuring that the resulting object really is a kernel, and proceed as follows: As mentioned above, there exists some Borel set $\Lambda \in \mathcal{B}([0,1])$ with $\lambda(\Lambda) = 1$ such that for every $x \in \Lambda$ we have $\delta'(x) = \hat{w}_\delta(x)$. Then defining $w_\delta : [0,1] \to [0,2]$ by

$$w_\delta(x) = \hat{w}_\delta(x)\mathbf{1}_\Lambda(x) + \frac{\delta(x)}{x}\mathbf{1}_{\Lambda^c}(x) = \delta'(x)\mathbf{1}_\Lambda(x) + \frac{\delta(x)}{x}\mathbf{1}_{\Lambda^c}(x).$$

obviously w_δ is Borel measurable and fulfills $w_\delta(x) \geq \frac{\delta(x)}{x}$ for every $x \in [0,1]$.

Theorem 2. *Let S_δ be a LSL copula with diagonal $\delta \in \mathcal{D}^{LSL}$ and let w_δ denote the measurable version of δ' as constructed above. Then (a version of) the Markov Kernel $K_{S_\delta}(\cdot,\cdot)$ of S_δ is given by*

$$K_{S_\delta}(x,[0,y]) = \begin{cases} \frac{y}{x}w_\delta(x) - \frac{y}{x^2}\delta(x) & \text{if } y < x, \\ \frac{1}{y}\delta(y) & \text{otherwise.} \end{cases} \tag{4}$$

Proof. For fixed $y \in [0,1]$ the mapping $x \mapsto K(x,[0,y])$ is obviously Borel-measurable, so measurability of $x \mapsto K(x,F)$ for arbitrary $F \in \mathcal{B}([0,1])$ follows via a standard Dynkin system argument. Now let $x \in [0,1]$ be arbitrary but fixed. Then the function $y \mapsto F_{S_\delta}(y) = K_{S_\delta}(x,[0,y])$ clearly is non-decreasing on $[0,x)$ and on $[x,1]$. Furthermore, distinguishing the two cases (i) $x \in \Lambda$ and (ii) $x \in \Lambda^c$ it is straightforward to verify that $F_{S_\delta}(x-) = w_\delta(x) - \frac{\delta(x)}{x} \leq \frac{\delta(x)}{x} = F_{S_\delta}(x)$. So altogether the function $y \mapsto F_{S_\delta}(y) = K_{S_\delta}(x,[0,y])$ is non-decreasing on $[0,1]$. Considering that F_{S_δ} is obviously right-continuous and that $F_{S_\delta}(0) = 0, F_{S_\delta}(1) = 1$ holds it follows that F_{S_δ} is a distribution function.
It remains to show that K_{S_δ} is a Markov kernel of S_δ, which can be done as follows: For $x \leq y$ it follows that

$$\int_{[0,x]} K_{S_\delta}(t,[0,y])\, d\lambda(t) = \int_{[0,x]} \frac{1}{y}\delta(y)\, d\lambda(t) = x\frac{\delta(y)}{y}.$$

Moreover, for $y < x$ we get

$$\begin{aligned} \int_{[0,x]} K_{S_\delta}(t,[0,y])\, d\lambda(t) &= \int_{[0,y]} K_{S_\delta}(t,[0,y])\, d\lambda(t) + \int_{[y,x]} K_{S_\delta}(t,[0,y])\, d\lambda(t) \\ &= \int_{[0,y]} \frac{1}{y}\delta(y)\, d\lambda(t) + \int_{[y,x]} \frac{y}{t}w_\delta(t) - \frac{y}{t^2}\delta(t)\, d\lambda(t) \\ &= \delta(y) + y\int_{[y,x]} \frac{tw_\delta(t) - \delta(t)}{t^2}\, d\lambda(t) \\ &= \delta(y) + y\left(\frac{\delta(x)}{x} - \frac{\delta(y)}{y}\right) = y\frac{\delta(x)}{x}. \end{aligned}$$

This altogether shows that $\int_{[0,x]} K_{S_\delta}(t,[0,y])\,d\lambda(t) = S_\delta(x,y)$, so $K_{S_\delta}(\cdot,\cdot)$ is indeed a Markov kernel of S_δ and the proof is complete. □

Consider a bivariate random vector (X,Y) with continuous distribution function H, marginals F and G, and let $(X_1,Y_1),(X_2,Y_2),(X_3,Y_3)$ be a sample of (X,Y). Then Kendall's τ of (X,Y) is given by

$$\tau_{X,Y} = \mathbb{P}((X_2-X_1)(Y_2-Y_1)>0) - \mathbb{P}((X_2-X_1)(Y_2-Y_1)>0) \tag{5}$$

and Spearman's ρ by

$$\rho_{X,Y} = 3\left(\mathbb{P}((X_2-X_1)(Y_3-Y_1)>0) - \mathbb{P}((X_2-X_1)(Y_3-Y_1)>0)\right). \tag{6}$$

For more background on concordance measures we refer to [7], for the relationship of ρ and τ to [9].

As a direct consequence of Thm. 2 the following simple formulas for Kendall's τ and Spearman's ρ of LSL copulas can be derived easily:

Lemma 1. *For every LSL copula $S_\delta \in \mathcal{C}^{LSL}$ the following formulas hold:*

$$\rho(S_\delta) = 12\int_{[0,1]} \delta(x)x\,d\lambda(x) - 3,$$
$$\tau(S_\delta) = 4\int_{[0,1]} \frac{\delta(x)^2}{x}\,d\lambda(x) - 1.$$

Example 1. *For $\theta\in[0,1]$ define the function $\delta\in\mathcal{D}^{LSL}$ by $\delta_\theta(t)=\theta t^2+(1-\theta)t$. Then the corresponding LSL copula S_{δ_θ} is a convex combination of Π and M. Applying Lemma 1 directly yields*

$$\tau(S_{\delta_\theta}) = \frac{\theta^2-4\theta}{3}+1,$$
$$\rho(S_{\delta_\theta}) = 1-\theta.$$

Considering $\theta_1=0.2,\theta_2=0.5,\theta_3=0.8$ the corresponding pairs $(\tau(S_{\delta_{\theta_i}}),\rho(S_{\delta_{\theta_i}}))$ are given by $(0.75,0.8),(0.42,0.5),(0.15,0.2)$. Some samples of the LSL copulas $S_{\delta_{\theta_i}}$ are depicted in Fig. 1.

4 Monotonicity and Singularity

Recall that a bivariate copula C is said to be (see [7] and [12])

- positively quadrant dependent (PQD) if $C(x,y)\geq\Pi(x,y)$ holds for all $(x,y)\in[0,1]^2$.
- left tail decreasing (LTD) if, for any $y\in[0,1]$, the mapping $(0,1)\to[0,1]$ given by $x\mapsto\frac{C(x,y)}{x}$ is non-increasing.

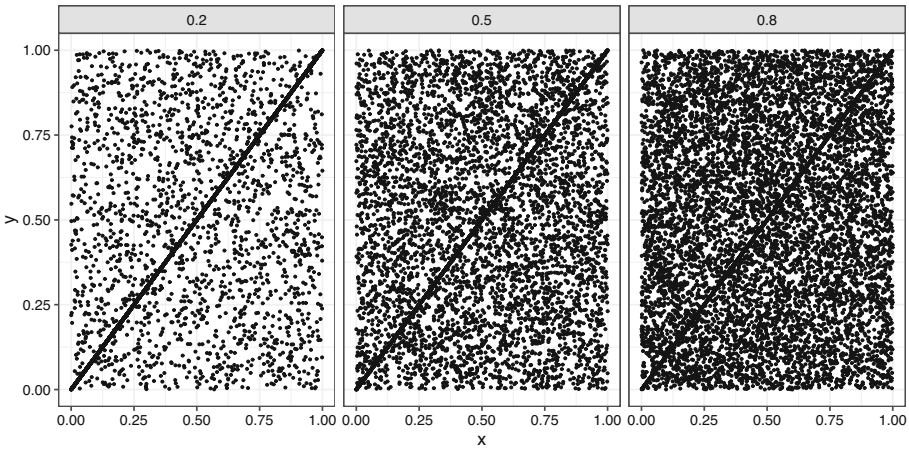


Fig. 1. Samples of size $n = 10.000$ of the LSL copulas $S_{\delta_{\theta_i}}$ with parameters $\theta_1 = 0.2, \theta_2 = 0.5, \theta_3 = 0.8$ as considered in Example 1.

– stochastically increasing (SI) if, for (a version of) the Markov kernel K_C and every $y \in (0,1)$ the mapping $x \mapsto K_C(x,[0,y])$ is non-increasing.

In [2] it was shown that $\Pi \leq S_\delta \leq M$ holds for all lower semilinear copulas $S_\delta \in \mathcal{C}^{LSL}$, so LSL copulas are PQD.
For arbitrary but fixed $y \in (0,1)$ the mapping

$$x \mapsto \frac{S_\delta(x,y)}{x} = \begin{cases} y\frac{\delta(x)}{x^2} & y \leq x \\ \frac{\delta(y)}{y} & y \geq x. \end{cases}$$

is non-increasing, implying that every LSL copula $S_\delta \in \mathcal{C}^{LSL}$ is LTD.
The following counterexample shows that LSL copulas are not necessarily SI.

Example 2. *Consider the points* $(0,0),(1/4,3/20),(1/2,5/12),(3/4,5/8),(1,1)$ *and let* $\delta : [0,1] \to [0,1]$ *denote the linear interpolation of these points. It is straightforward to verify that* $\delta \in \mathcal{D}^{LSL}$ *and* δ *fulfills the property that* $x \mapsto \frac{\delta(x)}{x}$ *is constant on two intervals. In this case it is straightforward to find some* $y \in (0,1)$ *such that the mapping* $x \mapsto K(x,[0,y])$ *is not non-increasing, so the corresponding LSL copula* S_δ *is not SI. Figure 2 depicts a sample of this very copula.*

We conclude this note by deriving a simple formula for the singular mass of LSL copulas.

Lemma 2. *For every LSL copula* S_δ *the singular mass is given by*

$$sing(S_\delta) = 2 \int_{[0,1]} \frac{\delta(x)}{x}\, d\lambda(x) - 1. \tag{7}$$

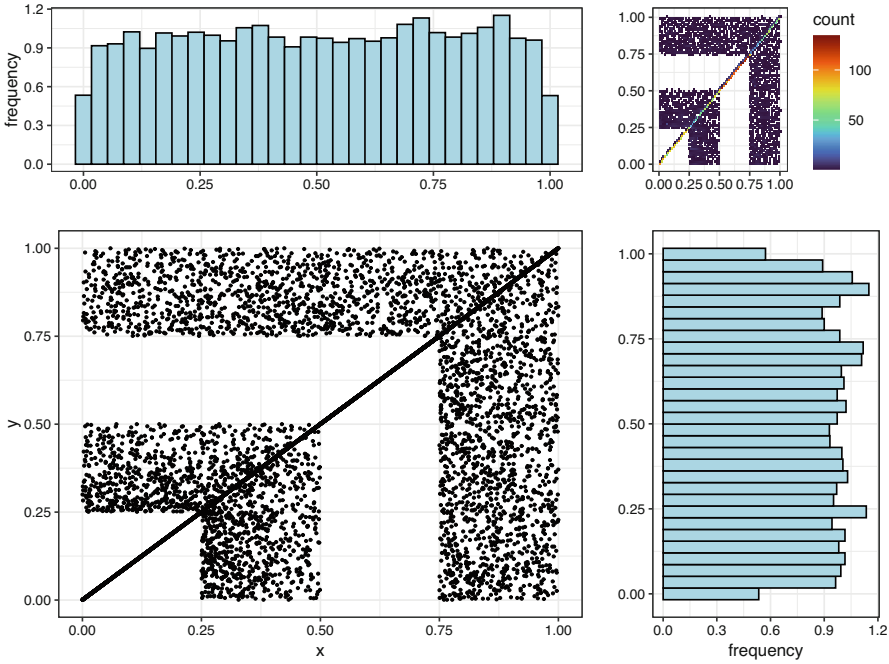


Fig. 2. Samples of size $n = 10.000$ of the LSL copula S_δ considered in Example 2 (lower felt panel); two-dimensional histogram (upper right panel) and marginal histograms (upper left and lower right panel).

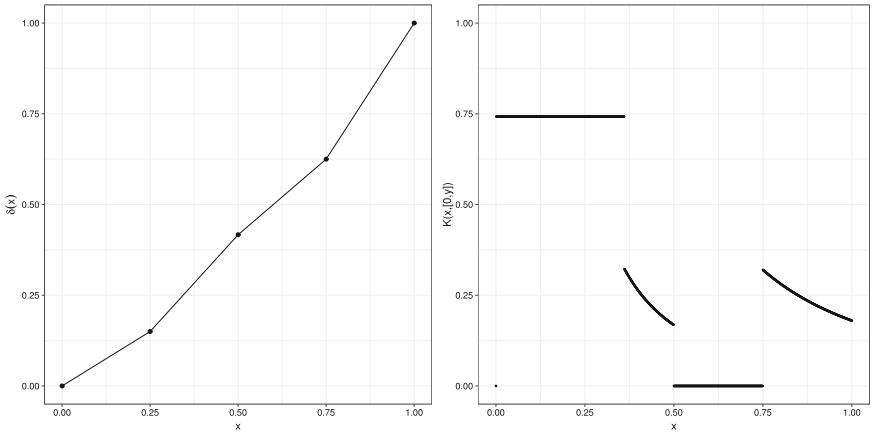


Fig. 3. The diagonal considered in Example 2 (left panel) and the mapping $x \mapsto K(x, [0, y])$ for $y = 0.36$ (right panel).

Proof. First of all notice that the only potential discontinuity of the conditional distribution function $y \mapsto K_{S_\delta}(x,[0,y])$ is the point x. Applying Thm. 2 and disintegration the stated identity therefore follows via

$$\begin{aligned} sing(S_\delta) &= \int_{[0,1]} K_{S_\delta}(x,\{x\})\, d\lambda(x) = \int_{[0,1]} K_{S_\delta}(x,[0,x]) - K_{S_\delta}(x,[0,x))\, d\lambda(x) \\ &= \int_{[0,1]} \frac{1}{x}\delta(x) - \frac{x}{x}w_\delta(x) + \frac{x}{x^2}\delta(x)\, d\lambda(x) = \int_{[0,1]} 2\frac{1}{x}\delta(x) - w_\delta(x)\, d\lambda(x) \\ &= 2\int_{[0,1]} \frac{1}{x}\delta(x)\, d\lambda(x) - \int_{[0,1]} w_\delta(x)\, d\lambda(x) \\ &= 2\int_{[0,1]} \frac{1}{x}\delta(x)\, d\lambda(x) - 1. \end{aligned}$$

□

Acknowledgements. The first author gratefully acknowledges the support of the EXDIGIT (Excellence in Digital Sciences and Interdisciplinary Technology) project, funded by Land Salzburg under grant number 20204-WISS/263/6-6022. The second author gratefully acknowledges the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

References

1. Arias-García, J.J., De Meyer, H., De Baets, B.: Multivariate upper semilinear copulas. Inform. Sci. **360**, 289–300 (2016)
2. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Semilinear copulas. Fuzzy Sets Syst. **159**, 63–76 (2008)
3. Durante, F., Sempi, C.: Principles of Copula Theory. Chapman and Hall/CRC (2015)
4. Durante, F., Fernández-Sánchez, J., Úbeda-Flores, M.: Extreme semilinear copulas. Fuzzy Sets Syst. **428**, 121–137 (2022)
5. Kallenberg, O.: Foundations of Modern Probability. Springer, New York (1997)
6. Kasper, T., Fuchs, S., Trutschnig, W.: On weak conditional convergence of bivariate archimedean and extreme value copulas, and consequences to nonparametric estimation. Bernoulli **27**(4), 2217–2240 (2021)
7. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)
8. Rudin, W.: Real and Complex Analysis. McGraw Hill International Editions, Singapore (1987)
9. Schreyer, M., Paulin, R., Trutschnig, W.: On the exact region determined by Kendall's tau and Spearman's rho. J. R. Stat. Soc. Ser. B Stat Methodol. **79**(2), 613–633 (2017)
10. Sloot, H., Scherer, M.: A probabilistic view on semilinear copulas. Inform. Sci. **512**, 258–276 (2020)
11. Trutschnig, W.: On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. **384**, 690–705 (2011)
12. Tschimpke, M., Fuchs, S.: Total positivity of copulas from a Markov kernel perspective. J. Math. Anal. Appl. **518**, 126629 (2023)

Addressing Incomplete Data in Two-Way Contingency Tables with Three-Level Variables

Pitchayanin Makapawee, Kanyawee Kamkongkaew, and Monchai Kooakachai(✉)

Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
Monchai.K@chula.ac.th

Abstract. Handling incomplete, imprecise, or missing observations in two-way contingency tables presents a challenge in accurately assigning uncertain data points to appropriate groups. Various approaches have been proposed to address this issue. Traditionally, the Expectation and Maximization (EM) algorithm has been utilized to estimate the group of incomplete observations. Subsequently, two alternative methods, derived from modifications to the formula used in the E-step of the algorithm, were introduced. However, these methods have mainly been evaluated in scenarios where categorical variables have two levels. This study aims to broaden the comparison of these methods to contingency tables with categorical variables featuring three levels. Through a simulation study, our findings indicate that while the classical EM algorithm generally performs well, there are specific parameter configurations where alternative methods demonstrate superior performance.

Keywords: Contingency tables · EM algorithm · Incomplete data

1 Introduction

When gathering data from a survey that includes questions about categorical variables, respondents may choose to skip certain questions due to discomfort or their perceived irrelevance. This leads to uncertainty about the level or group to which each observation belongs. As a result, the summarized data in the respective two-way contingency table is considered incomplete due to the lack or uncertainty of information regarding one or more categorical variables.

Addressing imprecise or uncertain observations in analysis poses a challenge. While researchers may opt to discard incomplete observations, doing so is not advisable for certain types of data as it can result in the loss of valuable information. Instead of discarding incomplete data points, statisticians have developed methods to effectively handle them within the analysis. One straightforward approach to address incomplete observations involves using maximum likelihood estimates (MLEs) of population proportions, derived from both completely and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 264–272, 2024.
https://doi.org/10.1007/978-3-031-65993-5_32

partially classified cases within the observed data. Ehlers [2] provided a comprehensive summary of MLE procedures for incomplete two-way contingency tables. Additionally, a general framework for maximizing likelihood under incomplete data is outlined in [1]. Beyond MLEs, fractional imputation methods for incomplete two-way contingency tables are also discussed in [3].

Countless efforts detailed in academic literature aim to improve the techniques used to fill in missing data in contingency tables that involve variables with two levels. For instance, Takai and Kano [8] broadened the original estimation outcomes to include scenarios involving a 2×2 contingency table with nonignorable responses. Furthermore, in 2010, Petitrenaud [5] introduced a technique to adjust uncertain data within a 2×2 contingency table using both frequentist principles and belief functions. Moreover, Yu, Bhadra et al. [9] introduced two tests based on both frequentist and Bayesian frameworks to evaluate independence within a 2×2 contingency table with random margins. These developments highlight the various efforts directed towards enhancing imputation techniques for incomplete two-way contingency tables.

A conventional approach to handling uncertain data in contingency tables is to utilize the Expectation and Maximization (EM) algorithm. While straightforward, one limitation of this classical EM algorithm is that, when replacing imprecise or uncertain data, it only utilizes observed data from the same level as the observation. To illustrate, consider two survey questions: favorite color and marital status. If individuals who identified as single did not specify their favorite color, their color preferences would be estimated based on responses from other single individuals. This restriction may be problematic if we assume independence between these two variables. In 2022, two distinct methods were introduced to address this gap by imputing incomplete data back into the original contingency table [4]. These novel approaches are founded on straightforward principles: one involves relaxing a constraint in the classical EM method and incorporating all available observed data, while the other utilizes the discrete uniform distribution. We will explore these approaches further in the following section. To determine the most effective approach across various parameter settings, simulations will be conducted. The simulation methodology proposed by Sulewski [7] for 2×2 contingency tables will serve as a foundation for our study. The methods introduced in [4] will be adapted and refined to operate within the 3×3 framework, thereby emerging as potential candidates for the optimal method.

2 Problem Description

For a given two-way contingency table with three-level variables, namely C_1 and C_2, define the true probability that an observation falls in cell (i, j) as π_{ij}, where $\pi_{ij} \geq 0$ and $\sum \pi_{ij} = 1$ so the complete data Y have a multinomial distribution with parameters n and $\pi = (\pi_{11}, \pi_{12}, \ldots, \pi_{33})$.

The estimation will be conducted utilizing the following outcomes. First, let $Y_{i+} = \sum_{j=1}^{3} Y_{ij}$ be the total count of row i and also $\pi_{i+} = \sum_{j=1}^{3} \pi_{ij}$ be the corresponding probability that an element falls in that particular level $C_1 = i$. If

Y_{i+} elements are independent, each having a probability distribution π_{ij}/π_{i+} for $j = 1, 2, 3$, then given the row total missing M and the vector of cell probabilities π, the elements of row i have a multinomial distribution:

$$Y_{i1}, Y_{i2}, Y_{i3} \mid M, \pi \sim \text{Multinomial}(m; \frac{\pi_{i1}}{\pi_{i+}}, \frac{\pi_{i2}}{\pi_{i+}}, \frac{\pi_{i3}}{\pi_{i+}}) \tag{1}$$

Observe that those probabilities in a parameter π was obtained from the maximum-likelihood estimation. Likewise, if we define notations in a consistent manner and hold the column totals fixed, then the elements of column j will follow a multinomial distribution:

$$Y_{1j}, Y_{2j}, Y_{3j} \mid M, \pi \sim \text{Multinomial}(m; \frac{\pi_{1j}}{\pi_{+j}}, \frac{\pi_{2j}}{\pi_{+j}}, \frac{\pi_{3j}}{\pi_{+j}}) \tag{2}$$

Now, let's contextualize everything within the framework of the problem. Suppose the data is sampled and uncertain values arise. The observed data can be partitioned into five parts denoted by A, B, C, D and E, respectively. Part A comprises observations with complete information on levels from both C_1 and C_2. Part B encompasses observations with complete information on C_1 but no information available on C_2. Conversely, part C contains observations completely classified for C_2 but lacking information on C_1. Additionally, part D includes observations that are fully observed on C_1 but only partially observed on C_2. Lastly, part E consists of observations that are fully observed on C_2 but only partially observed on C_1. The last four parts will be imputed back to the calculation based on the following three different algorithms.

2.1 Imputation by the EM Algorithm

The conventional approach to handle incomplete values involves the utilization of the EM algorithm. This methodology can be described as follows. Initially, let's assume that the complete data count, denoted as Y_{ij}, can be expressed as the aggregate of contributions from various components. Specifically,

$$Y_{ij} = Y_{ij}^A + Y_{ij}^B + Y_{ij}^C + Y_{ij}^D + Y_{ij}^E.$$

Part A is disregarded in the imputation process as the data is complete. To adjust the calculation of the sample in part B (where the categorical variable C_2 is entirely unknown), we first denote the total number of missing observations at level $C_1 = i$ as y_{im}^B. Then, according to Eq. (1), we can express

$$Y_{i1}^B, Y_{i2}^B, Y_{i3}^B \mid Y_{im}^B, \pi \sim \text{Multinomial}(y_{im}^B; \frac{\pi_{i1}}{\pi_{i+}}, \frac{\pi_{i2}}{\pi_{i+}}, \frac{\pi_{i3}}{\pi_{i+}}).$$

Note that the corresponding expected count is $E(Y_{ij}^B \mid Y_{im}^B = y_{im}^B, \pi) = y_{im}^B \left(\frac{\pi_{ij}}{\pi_{i+}}\right)$. Likewise, it follows from Eq. (2) that predictive distribution of

the missing data in part C, given the observed value and the vector of cell probabilities, is multinomial. In other words,

$$Y_{1j}^C, Y_{2j}^C, Y_{3j}^C \mid Y_{mj}^C, \pi \sim \text{Multinomial}(y_{mj}^C; \frac{\pi_{1j}}{\pi_{+j}}, \frac{\pi_{2j}}{\pi_{+j}}, \frac{\pi_{3j}}{\pi_{+j}}).$$

For a sample in part D, categorical variable C_1 is fully observed, while C_2 is only partially observed. For each $i = 1, 2, 3$ and $k = 1, 2, 3$, let $y_{i,(k)}^D$ denote the total number of imprecise observations belonging to $C_1 = i$, where the level for C_2 can be any value except $C_2 = k$. Similarly, define notations $\pi_{i,(k)}$ in the same manner, representing the corresponding probabilities. It can be deduced from Eq. (1) that the predictive distribution of the missing data in part D, given the observed value and the vector of cell probabilities, follows a multinomial distribution. In other words,

$$Y_{1,(k)}^D, Y_{2,(k)}^D, Y_{3,(k)}^D \mid Y_{i,(k)}^D, \pi \sim \text{Multinomial}(y_{i,(k)}^B; \frac{\pi_{i1}}{\pi_{i,(k)}}, \frac{\pi_{i2}}{\pi_{i,(k)}}, \frac{\pi_{i3}}{\pi_{i,(k)}}).$$

Finally, employing notations defined in a similar manner, the predictive distribution of the uncertain data in part E, given the observed value and the vector of cell probabilities, follows a multinomial distribution. In other words,

$$Y_{(k),1}^E, Y_{(k),2}^E, Y_{(k),3}^E \mid Y_{(k),j}^E, \pi \sim \text{Multinomial}(y_{(k),j}^E; \frac{\pi_{1j}}{\pi_{(k),j}}, \frac{\pi_{2j}}{\pi_{(k),j}}, \frac{\pi_{3j}}{\pi_{(k),j}}).$$

This distribution is derived from Eq. (2). Based on these predictive multinomial distributions, we can utilize their expectations in the calculation of the E-step in the EM algorithm. Specifically, let Y_{obs} represent the observed data and $\pi^{(r)}, r = 0, 1, 2, \ldots$ denote the rth step estimate of π. Then, in the E-step, the term $E(Y_{ij} \mid Y_{obs}, \pi^{(r)})$ is computed as follows:

$$E(Y_{ij} \mid Y_{obs}, \pi^{(r)}) = y_{ij}^A + y_{im}^B \left(\frac{\pi_{ij}^{(r)}}{\pi_{i+}^{(r)}}\right) + y_{mj}^C \left(\frac{\pi_{ij}^{(r)}}{\pi_{+j}^{(r)}}\right) + y_{i,(k)}^D \left(\frac{\pi_{ij}^{(r)}}{\pi_{i,(k)}^{(r)}}\right) + y_{(k),m}^E \left(\frac{\pi_{ij}^{(r)}}{\pi_{(k),j}^{(r)}}\right)$$

In the M-step, $\pi^{(r+1)}$ is updated by substituting the results from the E-step into the MLE of π for the complete data, as demonstrated in the equation below:

$$\pi_{ij}^{(r+1)} = \frac{1}{n}(E(Y_{ij} \mid Y_{obs}, \pi^{(r)}))$$

Here, n represents the total count of complete and incomplete data. This iterative process continues until convergence is achieved.

2.2 Modifications of Algorithms

Modification 1 (Adapted EM) : To illustrate the rationale behind the adapted EM method, we begin with an example. Let's consider the observed data in part C, where the level of C_1 is unknown but fully known in C_2. In the original EM algorithm, the estimated ratio is derived from complete data

restricted to the level of C_2 to which the observation belongs. Accordingly, the desired proportion takes the form $\pi_{ij}^{(r)}/\pi_{+j}^{(r)}$. However, if we eliminate this restriction, the estimated ratio can be computed from the entire set of complete data, not just the level of variable C_2. Therefore, the term $\pi_{ij}^{(r)}/\pi_{+j}^{(r)}$ should be replaced by $\sum_{j=1}^{3}\pi_{ij}^{(r)}/\sum_{j=1}^{3}\pi_{+j}^{(r)} = \pi_{i+}^{(r)}$. Similar principles can also be applied to the sample parts B, D, and E. Integrating all these aspects leads to the formulation of the new E-step in the algorithm as follows:

$$E(Y_{ij} \mid Y_{obs}, \pi^r) = y_{ij}^A + y_{im}^B \pi_{+j}^{(r)} + y_{mj}^C \pi_{i+}^{(r)} + y_{i,(k)}^D \left(\frac{\pi_{+j}^{(r)}}{\sum_{i=1}^{3} \pi_{i,(k)}^{(r)}} \right) + y_{(k),m}^E \left(\frac{\pi_{i+}^{(r)}}{\sum_{j=1}^{3} \pi_{(k),j}^{(r)}} \right)$$

We iterate through the E-step and M-step multiple times until the parameters π_{ij} converge.

Modification 2 (Uniform): The second method stems from the notion that, in certain scenarios, the true level of uncertain data should be entirely random, independent of available data. Hence, uncertain observations are factored into the calculation using the discrete uniform distribution. It's important to note that this method does not entail iterative algorithms. The estimation is derived from the following equation:

$$E(Y_{ij} \mid Y_{obs}) = y_{ij}^A + y_{im}^B \left(\frac{1}{3}\right) + y_{mj}^C \left(\frac{1}{3}\right) + y_{i,(k)}^D \left(\frac{1}{2}\right) + y_{(k),m}^E \left(\frac{1}{2}\right)$$

Observe that the proportions in the last two terms are 0.5, reflecting partial information indicating that the observation will not be in one specific group. Once the expectation is obtained, the parameters π_{ij} are calculated using the standard MLE formula y_{ij}/n.

3 Simulation Study

The numerical investigation is divided into two segments. Initially, simulations are conducted using a particular type of symmetric contingency tables, wherein the performance of each method is compared. Conversely, the second segment describes a simulation study conducted without imposing any restrictions on the contingency table.

3.1 Case I: Symmetric Contingency Tables

In a contingency table, we represent the entry corresponding to $C_1 = i$ and $C_2 = j$ as a_{ij}. A contingency table is considered symmetric if $a_{ij} = a_{ji}$ for all indices i and j. Notably, an unconditioned contingency table featuring three levels for each variable involves a collective of 9 parameters. However, leveraging symmetry, we can establish $a := \pi_{12} = \pi_{21}$, $b := \pi_{13} = \pi_{31}$, $c := \pi_{23} = \pi_{32}$, and $k := \pi_{11} = \pi_{22} = \pi_{33}$, thereby reducing the number of parameters to

four. Careful selection of simulated values for a, b, c, and k is crucial to ensure comprehensive coverage of potential scenarios. The constraints $2a+2b+2c+3k = 1$ and $a, b, c, k > 0$ narrow down the choices to only three parameters. To cover a diverse range of scenarios, a, b, and c were selected from the range of 0.03–0.15, incrementing by 0.03. To assess the performance of the three algorithms, we follow these steps:

Step 1 Let p be the probability of drawing a complete observation. In that case, the levels of C_1 and C_2 are known. Conversely, an incomplete observation (with probability $q = 1 - p$) occurs when one or more categorical variables' levels are either unknown or partially known. If an incomplete observation is selected, it will be uniformly allocated to sample parts B, C, D, or E. To cover various scenarios, we'll use five values of p: $4/5, 2/3, 1/2, 1/3$, and $1/5$.

Step 2 Selecting one set of a, b, c, and p values from a possible 625 choices, we utilize R programming to draw a sample of size 1000 from a multinomial distribution with parameters $n = 1000$ and $\pi = (\pi_{11}, \pi_{12}, \ldots, \pi_{33})$. After completing the random sampling, we tabulate the results.

Step 3 Using the tabulated results from step 2, we proceed to estimate the parameters $\pi_{11}, \pi_{12}, \ldots, \pi_{33}$ using three different algorithms outlined in the preceding section. For both the classical EM algorithm and the adapted EM algorithm, we will iteratively run them at least 100 times or until the parameters converge.

Step 4 For every set of parameters, repeat steps 3 and 4 for 500 iterations, then calculate the root mean square error for each parameter estimate.

Figure 1 shows two 3D scatterplots where each point represents the method yielding the minimum average root mean square error in the estimation for $\pi_{11}, \pi_{12}, \ldots, \pi_{33}$. While the figure generally indicates that the classical EM tends to perform well, its presentation may be somewhat difficult to read, and it may not provide insight into situations where alternative methods perform better. Therefore, we provide additional details here. When $p = 4/5$, there are 3 cases where the Adapted EM method performs best, which are characterized by the tuples $(a, b, c) = (.06, .15, .12)$, $(.09, .09, .12)$, and $(.15, .12, .09)$. Also, there are 6 scenarios where the uniform approach performs best, characterized by the tuples $(a, b, c) = (.09, .09, .15)$, $(.09, .12, .12)$, $(.09, .15, .09)$, $(.09, .15, .12)$, $(.12, .09, .15)$ and $(.12, .12, .12)$. When $p = 2/3$, there are 10 instances where the Adapted EM performs best: $(a, b, c) = (.06, .12, .12)$, $(.06, .15, .12)$, $(.09, .09, .12)$, $(.09, .12, .12)$, $(.12, .09, .12)$, $(.12, .12, .09)$, $(.12, .12, .12)$, $(.12, .15, .09)$, $(.15, .09, .09)$ and $(.15, .12, .09)$. There are 4 scenarios where the uniform approach performs best, characterized by the tuples $(a, b, c) = (.09, .09, .15)$, $(.09, .12, .15)$, $(.09, .15, .12)$ and $(.12, .09, .15)$.

3.2 Case II: Unconditioned Contingency Tables

We replicate the simulation described in Case I, but now using unconditioned contingency tables. The total number of parameters amounts to 9, denoted as

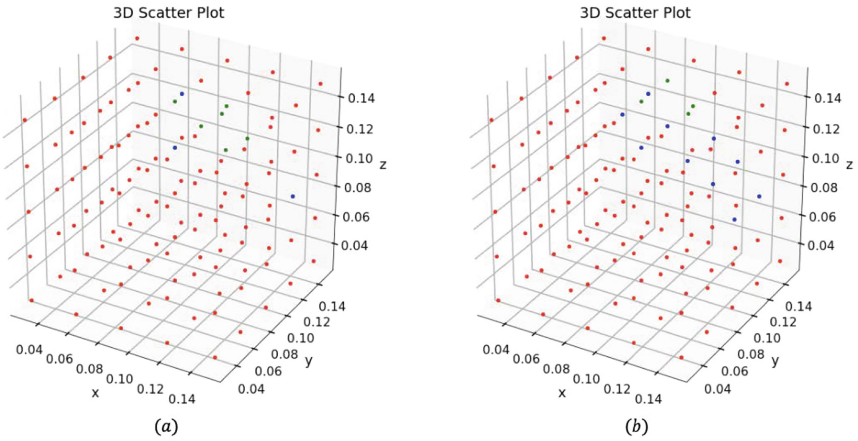


Fig. 1. 3D scatterplots where each point represents the method yielding the minimum average root mean square error in the estimation for $\pi_{11}, \pi_{12}, \ldots, \pi_{33}$. Axis labels refer to parameters as $x = a, y = b$, and $z = c$. Figures (a) and (b) refer to parameters $p = 4/5$ and $2/3$, respectively. The red, blue, and green points correspond to the classical EM, adapted EM, and uniform distribution approaches, respectively.

$\pi_{11}, \pi_{12}, \ldots, \pi_{33}$. These parameters are subject to constraints such that $\sum \pi_{ij} = 1$ and $\pi_{ij} > 0$. To capture a broad spectrum of scenarios, we generate potential values of π_{ij} 10,000 times, mirroring the approach used in the 10,000 example parameter cases. Subsequently, we apply the same four-step methodology as outlined in the previous case to compare the performance of the three algorithms. Evaluation of performance is based on the comparison of average root mean square errors. The results of these simulations are shown in Table 1, illustrating the percentage at which each method demonstrates superior performance across 10,000 simulations, with variations in the proportion of complete data p.

Table 1. Percentage of the best method across 10,000 simulations

p	(classic EM, adapted EM, uniform)
4/5	(96.8, 3.2, 0.0)
2/3	(94.8, 5.2, 0.0)
1/2	(92.6, 7.4, 0.0)
1/3	(89.5, 10.5, 0.0)
1/5	(90.7, 9.3, 0.0)

4 Conclusions and Discussion

In this study, we proposed and evaluated three algorithms designed to address incomplete observations in contingency tables, focusing on scenarios involving two categorical variables with three levels each. Our numerical investigation was divided into two distinct cases: symmetric and unconditioned. The key findings can be summarized as follows: In the symmetric case, the classical EM algorithm generally outperforms, yet there are specific parameter configurations where the alternative methods exhibit superior performance, although the precise pattern remains elusive.

Concerning the issue of sample size, all simulations were carried out with a sample size of 1000, as specified in Step 2 of Sect. 3.1. However, in numerous fields where contingency table analysis is relevant, the sample size could be considerably smaller. A comprehensive explanation of contingency table analysis can be found in [6]. Thus, as suggested by the source, we repeated the simulation with a probability of drawing a complete observation of $p = 4/5$, and conducted additional simulations with sample sizes reduced by factors of 10 and 5 (resulting in $n = 100$ and $n = 200$, respectively). The resulting outcomes closely resemble those shown in Fig. 1(a) where the classical EM algorithm tends to dominate in most cases. Nonetheless, a more thorough investigation is necessary to fully analyze the implications of using such smaller sample sizes.

In the unconditioned case, the classical EM algorithm retains its dominance, but we noted that the adapted EM algorithm tends to perform better as the probability of drawing a complete observation (p) decreases. Additionally, we did not observe any instances where the algorithm involving the uniform distribution demonstrated the best performance.

Acknowledgements. The author expresses gratitude to the editor, associate editor, and referees for their valuable feedback and suggestions, which significantly enhanced the quality of this article.

References

1. Couso, I., Dubois, D.: A general framework for maximizing likelihood under incomplete data. Int. J. Approx. Reason. **93**, 238–260 (2018)
2. Ehlers, R.: Maximum Likelihood Estimation Procedures for Categorical Data. University of Pretoria, Diss (2005)
3. Kang, S., Koehler, K., Larsen, M.: Fractional imputation for incomplete two-way contingency tables. Metrika **75**, 581–599 (2012)
4. Kooakachai, M.: Algorithms for the test of independence of two categorical variables over Uncertain Data. Thai J. Math. 62–74 (2022)
5. Petitrenaud, S.: Independence tests for uncertain data with a frequentist method. In: Borgelt, C., et al. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis. Advances in Intelligent and Soft Computing, vol. 77, pp. 519–526. Springer, Berlin, Heidelberg (2010)
6. Rudas, T.: Lectures on Categorical Data Analysis. Springer, US (2018)

7. Sulewski, P.: Some contributions to practice of 2× 2 contingency tables. J. Appl. Stat. **46**(8), 1438–1455 (2019)
8. Takai, K., Kano, Y.: Test of independence in a 2 × 2 contingency table with non-ignorable nonresponse via constrained EM algorithm. Comput. Stat. Data Anal. **52**(12), 5229–5241 (2008)
9. Yu, Y., Bhadra, D., Nandram, B.: Tests of independence for a 2× 2 Contingency Table with Random Margins. Int. J. Stat. Prob. **6**(2), 106–121 (2017)

Exploring Deep Reinforcement Learning Algorithms for Enhanced HVAC Control

Antonio Manjavacas(✉), Alejandro Campoy-Nieves, Miguel Molina-Solana, and Juan Gómez-Romero

Department of Computer Science and Artificial Intelligence, Universidad de Granada, Periodista Daniel Saucedo Aranda, 18071, Granada, Spain
manjavacas@ugr.es

Abstract. Heating, Ventilation, and Air Conditioning (HVAC) systems are one of the major sources of energy consumption in buildings. Typically, HVAC control has relied on reactive controllers, which often lack the ability to adapt to specific building dynamics. In recent years, Deep Reinforcement Learning (DRL) algorithms have emerged as a potential alternative to reactive controllers. However, these solutions are still immature, with a lack of standardisation and difficulties in real-world deployment. This paper presents an empirical evaluation of several state-of-the-art DRL algorithms for HVAC control, highlighting their main strengths and limitations. We emphasize the importance of using standard frameworks for comparative analysis, enabling a more comprehensive assessment of these innovative approaches to HVAC control.

Keywords: Reinforcement learning · Building energy optimization · HVAC

1 Introduction

The building and construction sectors are responsible for approximately 17% of the direct global CO_2 emissions, and over one third of the global energy consumption [8]. Specifically, Heating, Ventilation, and Air Conditioning (HVAC) systems account for 50% of the total energy demand in residential and non-residential buildings, making them one of the main sources of energy consumption [12].

HVAC power demand is often increased due to poor performance and imprecise control strategies [10]. Industry's HVAC solutions have typically involved rule-based controllers (RBCs) combined with Proportional-Integral-Derivative (PID) control, which, although simple and efficient, have limited scalability and adaptability to specific building conditions [15].

Alternatively, more sophisticated control methods based on building models have been explored. Such model predictive controllers (MPCs) rely on physical models to simulate the thermal dynamics of the building and analytically derive

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 273–280, 2024.
https://doi.org/10.1007/978-3-031-65993-5_33

optimal HVAC control [18]. However, developing and solving a detailed model for each building component is a complex task which requires expert knowledge and is not easily scalable. These limitations make implementation challenging, which has resulted in limited use in real-world scenarios [15].

Given the limitations of RBCs/PIDs, and the complexity of MPCs, in recent years there has been a particular trend towards control by (Deep) Reinforcement Learning (DRL) algorithms [11]. These methods can provide more sophisticated control than RBC/PID and typically require less complex building models than those required by MPC [5]. In turn, the main source of data for DRL controllers comes from the agent's interaction with a simulated environment, so the computational effort takes place offline —as opposed to online in the case of MPC—. Other benefits such as the ability to incorporate forecasting or the predictive capabilities of DRL are worth noting.

However, we find that several limitations, such as poor data efficiency, the need for —albeit simplified— building models, or hyperparameterisation, hinder the adoption of DRL in real-world scenarios. This has also motivated the study of hybrid approaches combining DRL with conventional control methods [1].

Thus, despite its potential, DRL for HVAC control is still a developing field. In general, the DRL methods proposed in the literature are ad hoc solutions, hardly comparable or transferable to other problems [4]. Contrary to other fields of machine learning, there is a lack of standardisation that needs to be addressed [16], and which should focus on providing a set of reference benchmarks on which to compare results and extrapolate solutions.

To fill this gap, in this paper we present an evaluation of several state-of-the-art DRL algorithms for HVAC control. We formulate a multi-objective optimization problem that aims to maximize comfort while minimizing energy consumption. Then, using the Sinergym framework [9], we evaluate the performance of three DRL agents against standard reactive controllers in six simulated scenarios, considering different buildings and weather conditions. The results allow us to identify the most promising algorithms, as well as their main limitations.

Thus, the rest of the paper is organised as follows. Section 2 introduces the HVAC control problem, followed by Sect. 3, where the methodology followed in the experimentation is described. In Sect. 4 we discuss the results obtained, leading to the conclusions and future work in Sect. 5.

2 Theoretical Background

2.1 Deep Reinforcement Learning

A reinforcement learning problem can be represented as a Markov Decision Process (MDP) where an agent interacts with its environment over a discrete sequence of time steps $\mathcal{T} = \{0, 1, 2, ...\}$. At each time step t, the agent performs an action $a_t \in \mathcal{A}$ based on the current state of the environment $s_t \in \mathcal{S}$. This action is given by a policy function π, where $a_t \sim \pi(\cdot|s_t)$. The action results in a new state $s_{t+1} \in \mathcal{S}$ and a reward signal $r_t \in \mathbb{R}$ representing the goodness of the action performed or the new state reached.

The objective of the agent is to learn an optimal policy π^* that maximizes the sum of discounted rewards over time, that is, $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, with $\gamma \in [0, 1]$. However, in those problems where $\mathcal{S}$ or $\mathcal{A}$ are too large —or even infinite—, we assume a parameterized policy π_θ, e.g., a neural network with weights θ, whose expected return is progressively approximated to that of an optimal policy by adjusting θ.

Combining reinforcement learning with neural networks gives rise to deep reinforcement learning algorithms such as Proximal Policy Optimisation (PPO) [14], Twin Delayed DDPG (TD3) [6] or Soft Actor-Critic (SAC) [7]. Each algorithm has a number of peculiarities in the way it approximates the optimal policy.

2.2 Formulation of the HVAC Control Problem

Recent literature reviews have previously explored the application of DRL algorithms in building HVAC control, such as: [17,19]. These studies summarize the variety of applications, perspectives and objectives addressed in this domain.

In this case, we formulate a multi-objective optimisation problem involving the continuous control of the cooling and heating setpoints of an HVAC system. The agent must learn to minimise the energy consumption of the building while ensuring the thermal comfort of its occupants.

This is a Partially Observable MDP (POMDP), where the actions taken by the agent are composed of continuous values representing the cooling and heating setpoints of the n controlled zones, that is, $a_t \in \mathcal{A}$, with $a_t \in \mathbb{R}^{2n}$. The agent's observations, $o_t \in \mathcal{O} \subset \mathcal{S}$, contain representative information about the current state of the building, such as indoor and outdoor temperatures, power consumption, occupancy, humidity, solar radiation, etc. An unknown transition function $f : \{\mathcal{S}, \mathcal{A}\} \rightarrow \mathcal{S}$ emulates the building dynamics after applying a_t over s_t, leading to s_{t+1}. In our case, these dynamics are provided and computed by a building simulator.

The employed reward function is a weighted sum of comfort and consumption terms, defined as follows:

$$R = -\omega \ \lambda_P \ P_t - (1 - \omega) \ \lambda_T \ e^{max(T_t - T_{low}, 0) + max(T_t - T_{up}, 0)} \tag{1}$$

where: ω is a constant weighing the importance assigned to comfort and consumption; P_t is the current building power consumption (in W); T_t (in °C) is the indoor temperature of the building; T_{low} and T_{up} are the assumed comfort range limits, and λ_P and λ_T are two scaling factors used to balance consumption and temperature magnitudes, respectively.

Note how the temperature-related component of the reward grows exponentially as the indoor temperature moves away from the comfort range. On the other hand, the reward is expressed in negative terms as we seek to maximise it.

3 Methodology

Our aim is to evaluate the performance of different DRL algorithms in HVAC control, given a series of standard building environments. We use the Sinergym framework to facilitate the execution and reproducibility of these experiments. In the following subsections, we detail the environments, agents and additional settings considered for this experimentation.

3.1 Environments

The following building models included in Sinergym where considered:

- *5ZoneAutoDXVAV*. A single-story office building divided into 5 zones. Its floor area is 463.6 m^2, and it is equipped with a VAV package (DX cooling coil and gas heating coils) with fully auto-sized input. The HVAC control corresponds to the central cooling and heating setpoints of the building.
- *2ZoneDatacenter*. A 491.3 m^2 datacenter divided in two zones. Each zone has an HVAC system consisting of air economizers, evaporative coolers, DX cooling coil, chilled water coil, and VAV. The main source of heat comes from the hosted servers. In this case, the cooling and heating setpoints for both west and east zones are controlled.

Each building is simulated with EnergyPlus under three different weather conditions, leading to a total of six simulated environments. The corresponding weather datasets and are as follows:

- *Hot dry*. Climate corresponding to Arizona (USA), with an average annual temperature of 21.7°C, and an average annual relative humidity of 34.9%.
- *Mixed humid*. Climate of New York (USA), with an average annual temperature of 12.6°C, and an average annual relative humidity of 68.5%
- *Cool marine*. Corresponds to Washington (USA). The average annual temperature is 9.3°C, while the average annual relative humidity is 81.1%.

3.2 DRL Algorithms and Configuration

The DRL algorithms used were PPO, TD3 and SAC. We used the algorithms implementations available in Stable-Baselines3 [13], and the hyperparameters shown in Table 1. These agents were trained for 40 episodes with evaluations every 10 episodes, being each episode equivalent to one simulated year of the building.

Subsequently, we used the results returned by the best models to compare performance, both between them and with the default RBC controllers of each building.

Table 1. Hyperparameters chosen for PPO, TD3 and SAC

PPO		**TD3**		**SAC**	
learning_rate	0.001	**learning_rate**	0.003	**learning_rate**	0.0003
n_steps	4096	**buffer_size**	1e6	**buffer_size**	1e6
batch_size	128	**batch_size**	100	**batch_size**	256
n_epochs	15	**tau**	0.005	**tau**	0.005
gamma	0.9	**gamma**	0.9	**gamma**	0.99
gae_lambda	0.9				

3.3 Experimental Settings

We consider the default observation and action spaces for the Sinergym buildings introduced in Sect. 3.1. Both actions and observations values are normalized using Sinergym wrappers. Actions are performed every 15 simulated minutes, which corresponds to 4 timesteps per hour. Furthermore, in order to improve generalization and reduce overfitting, random noise is added to the weather dataset in each episode [9].

The reward function used is the *ExpReward* implemented by default in Sinergym, similar to the one described in Eq. 1. The comfort ranges are those recommended by ASHRAE for residential buildings ($[23, 26]\,°C$ from June to September and $[20, 23.5]\,°C$ for the rest of the year) [2] and datacenters ($[18, 27]\,°C$) [3]. Comfort and power consumption are given equal weight, with $\omega = 0.5$ in Eq. 1.

All the experiments were conducted using a computer with an Intel Core i7-11800 H 2.30 GHz, NVIDIA GeForce RTX 3060 Mobile, and 16 GB of RAM.

4 Results

Once the DRL agents have been trained, we proceed to their evaluation. Figure 1 shows the results obtained for 5ZoneAutoDXVAV. SAC is the best performing DRL agent in terms of mean reward per episode, followed by TD3 and PPO. We observe that RBC outperforms the DRL controllers in terms of power demand, while all manage to obtain close results regarding comfort violation.

Note how variations with respect the comfort ranges are minimal in all cases, with comfort violations of less than $\pm 1\,°C$. On this basis, we consider that this building does not present a challenge in terms of temperature control, with demand being the main differentiating factor between the proposed controllers.

On the other hand, Fig. 2 illustrates that, in the case of 2ZoneDatacenter, DRL agents manage to outperform RBC in terms of reward. Note how, in most of the cases, DRL agents provide better comfort guarantees while higher energy consumption.

Based on these results, we observe that when it comes to more complex buildings —i.e., with a greater number of controlled zones— DRL agents manage

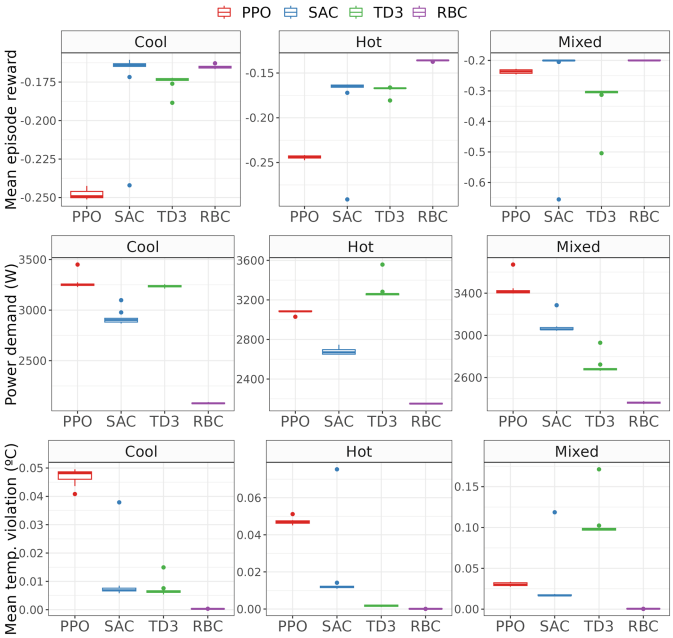


Fig. 1. Results after 10 evaluation episodes in 5ZoneAutoDXVAV

to perform better than an ad hoc RBC. Results also show how comfort is more easily achieved in colder climates, with better rewards for all controllers in both buildings.

5 Conclusions and Future Work

In this paper we have evaluated different DRL agents for HVAC control. Using the Sinergym framework and following a standardised approach, we have compared the performance of three state-of-the-art algorithms in six different environments in terms of comfort and energy consumption, both among themselves and against reactive controllers.

The results show that DRL controllers are able to compete with RBCs without extensive knowledge of the controlled environments. Overall, DRL agents performed similarly to RBCs in terms of rewards, achieving improved comfort but higher energy consumption. To further improve performance, we suggest including additional energy-related variables in the agents' observations.

We plan to conduct a more extensive experimentation involving a larger number of buildings and more complex systems, where the differences between methods may be more pronounced. However, we believe that the results obtained can serve as an important reference for future experimentation and demonstrates the potential of DRL agents in building energy optimisation.

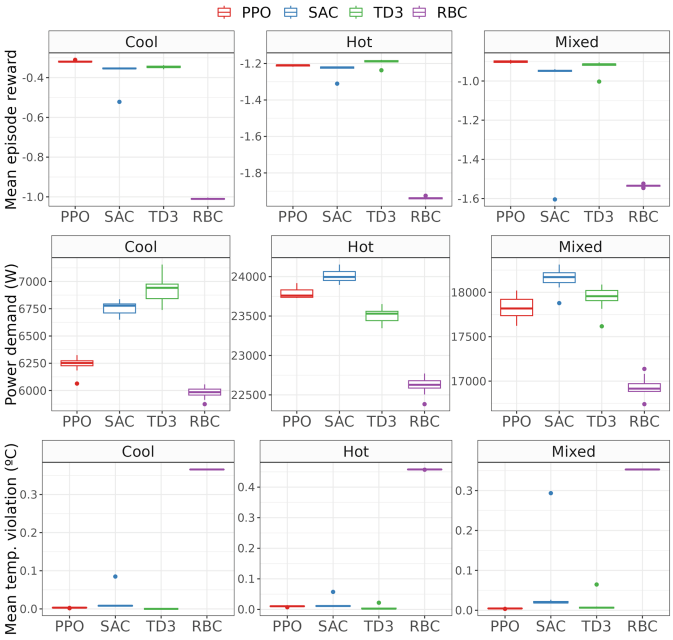


Fig. 2. Results after 10 evaluation episodes in 2ZoneDatacenter

As future work, we also plan to assess the generalisability of these agents by exploring advanced training techniques that facilitate their deployment in unseen environments. In addition, we intend to deploy these agents in real buildings, addressing the common sim-to-real problem in reinforcement learning.

Finally, we encourage the use of standard frameworks such as Sinergym to provide reproducible and comparable results. Data sharing, public models, and reproducible evaluations are some of the factors that can contribute to the maturity and development of building energy optimisation.

Acknowledgements. Acknowledgements to grants SINERGY (PID2021-125537NA-I00) funded by MICIU/AEI/10.13039/501100011033 and ERDF, IA4TES (MIA.2021.M04.0008) funded by the Spanish Ministry of Economic Affairs and Digital Transformation (NextGenerationEU), D3S (P21.00247) funded by ERDF/Junta de Andalucía, and SE2021 UGR IFMIF-DONES.

References

1. Arroyo, J., Manna, C., Spiessens, F., Helsen, L.: Reinforced model predictive control (RL-MPC) for building energy management. Appl. Energy **309**, 118346 (2022)
2. ASHRAE 55-2004: Thermal Environmental Conditions for Human Occupancy. ASHRAE (2004)
3. ASHRAE TC9.9: Data Center Power Equipment Thermal Guidelines and Best Practices systems. ASHRAE (2016)

4. Biemann, M., Scheller, F., Liu, X., Huang, L.: Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. Appl. Energy **298**, 117164 (2021)
5. Deng, X., Zhang, Y., Qi, H.: Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning. Build. Environ. **211**, 108680 (2022)
6. Fujimoto, S., Hoof, H., Meger, D.: Addressing function approximation error in actor-critic methods. In: Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 1587–1596 (2018)
7. Haarnoja, T., Zhou, A., Abbeel, P, Levine, S.: Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 1861–1870 (2018)
8. International Energy Agency Tracking Report on Buildings. (IEA, 2021). Available at https://www.iea.org/reports/tracking-buildings-2021
9. Jiménez-Raboso, J., Campoy-Nieves, A., Manjavacas-Lucas, A., Gómez-Romero, J., Molina-Solana, M.: Sinergym: a building simulation and control framework for training reinforcement learning agents. In: Proceedings of the 8th ACM International Conference On Systems For Energy-Efficient Buildings, Cities, And Transportation, pp. 319–323 (2021)
10. Mawson, V., Hughes, B.: Optimisation of HVAC control and manufacturing schedules for the reduction of peak energy demand in the manufacturing sector. Energy **227**(3), 120436 (2021)
11. Nagy, Z., Henze, G., Dey, S., Arroyo, J., Helsen, L., Zhang, X., et al.: Ten questions concerning reinforcement learning for building energy management. Build. Environ. **241**, 110435 (2023)
12. Pérez-Lombard, L., Ortiz, J., Pout, C.: A review on buildings energy consumption information. Energy Build. **40**, 394–398 (2008)
13. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N.: Stable-Baselines3: reliable reinforcement learning implementations. J. Mach. Learn. Res. **22**(268), 1–8 (2021)
14. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. ArXiv Preprint arXiv:1707.06347 (2017)
15. Serale, G., Fiorentini, M., Capozzoli, A., Bernardini, D., Bemporad, A.: Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities. Energies **11**, 631 (2018)
16. Vázquez-Canteli, J., Nagy, Z.: Reinforcement learning for demand response: a review of algorithms and modeling techniques. Appl. Energy **23**, 1072–1089 (2019)
17. Yang, T., Zhao, L., Li, W., Zomaya, A.: Reinforcement learning in sustainable energy and electric systems: a survey. Annu. Rev. Control. **49**, 145–163 (2020)
18. Yao, Y., Shekhar, D.: State of the art review on model predictive control (MPC) in heating ventilation and air-conditioning (HVAC) field. Build. Environ. **200**, 107952 (2021)
19. Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., Guan, X.: A review of deep reinforcement learning for smart building energy management. IEEE Internet Things J. **8**(15), 12046–12063 (2021)

Clustering on the d-Dimensional Sphere: Latest Trends and the Role of Poisson Kernel-Based Density Models

Marianthi Markatou[1,2(✉)]

[1] Department of Biostatistics, SPHHP, University at Buffalo, The State University of New York, Buffalo, USA
[2] Jacobs School of Medicine and Biomedical Sciences, Buffalo, USA
markatou@buffalo.edu

Abstract. Scientific work produces many data sets that can be analyzed as unit vectors on a d-dimensional sphere. Specific examples include clustering of documents, the study of comets, and data that arise in the analysis of microarray experiments where interest centers in identifying groups of genes that are functionally related. We offer a succinct review of the literature on clustering on the d-dimensional sphere. We then discuss the role that Poisson kernel-based density models play in clustering, highlighting their connection with the Brownian motion and other classes of densities. A special class of kernels, called diffusion kernels, is also discussed. Software for the use of a clustering method based on mixtures of Poisson kernel-based densities is also reviewed.

Keywords: Brownian motion · Clustering methods · D-dimensional sphere · Diffusion kernels · Poisson kernel-based densities

1 Introduction

Many scientific fields such as meteorology, biology, astronomy, natural language processing, produce data that can be thought of and analyzed as spherical data. For example, in directional statistics data reside on the unit hypersphere, defined as $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : ||x||^2 = 1\}$ of $\mathbb{R}^d$. Other examples include data sets where the magnitude of the measurement of variables is not the primary interest, but the direction is.

There is a sizeable literature on inference on the parameters of models that are used to describe spherical data. See, for example, Mardia and Jupp [16] and a recent paper by Dey and Jana [4]. Relatively less is written with respect to clustering. The literature includes clustering methods that can be categorized as both, parametric and nonparametric. Nonparametric approaches include the k-means algorithm proposed by Maitra and Ramler [15], spherical k-means proposed by Dhillon and Modha [5] and online spherical k-means proposed by Zhong [20]. Furthermore, studies on clustering spherical data based on kernel

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 281–285, 2024.
https://doi.org/10.1007/978-3-031-65993-5_34

densities have appeared in the literature (Di Marzio et al. [6]). A recently published paper [8] discusses a model for clustering spherical data using multidimensional scaling and geodesic distance. For details, see Fernando Vera et al. [8].

Parametric approaches use probability models to perform clustering. The literature includes the clustering algorithm of Banerjee et al. [1], where a finite mixture of von Mises-Fisher distributions is employed to cluster the data. A clustering algorithm based on mixtures of Watson distributions is also proposed, see for example Bijral et al. [3]. To model nonsymmetric cluster shapes Kim and SenGupta [14] proposed the use of Kent distributions [12]. This distribution is suitable for modeling elongated clusters. The authors also proposed spherical generalizations of two asymmetric circular distributions, generalized von Mises-Fisher for circular data and the three parameter generalized von Mises density with location parameter $\mu, k_1 > 0$ being the concentration parameter, and $k_2 \in [-1, 1]$ is the skewness parameter (Kim and SenGupta [13,14]).

An alternative, model-based, clustering algorithm has been proposed by Golzy and Markatou [10]. This algorithm is based on Poisson kernel-based densities. The algorithm outperforms existing popular algorithms in the case of poor separation between the different clusters, and the case of existence of increased noise in the data. Golzy and Markatou [10] discuss the construction of an EM-type algorithm based on mixtures of Poisson kernel-based densities, and provide results on clustering several data sets using the aforementioned model and algorithm.

2 Poisson Kernel-Based Density Models

Poisson kernel-based density models are based on a special diffusion kernel, the Poisson kernel. Despite its extensive role in the Dirichlet problem this kernel is not well known in the statistical literature. Ding et al. [7] discuss tests for uniformity on the d-dimensional sphere that are based on the Poisson kernel. We begin by discussing the concept of diffusion kernels.

3 Diffusion Kernels

In the field of machine learning kernels are used to provide a similarity metric on the input space $\mathcal{X}$, and must satisfy the requirements of being positive semidefinite and symmetric. In mathematics, diffusion kernels are defined as solutions of appropriate heat equations. A highly prominent example of a diffusion kernel is the normal kernel. The Poisson kernel, defined below, generalizes the key features of the Gaussian kernel to the d-dimensional sphere.

Ding et al. [7] define diffusion kernels via the desired mathematical properties that such a kernel must exhibit.

Definition 1. [7]. A family of symmetric kernels $K_t(r, s)$ defined on $\mathcal{X} \times \mathcal{X}$, where $\mathcal{X}$ is a sample space, with parameter $t \in (0, \infty)$ is called a **diffusion kernel family** with respect to a measure $u(\bullet)$ if the following properties hold.

1. The kernel is such that $K_t(r,s) \geq 0$.
2. The kernel satisfies the equation

$$K_{t_1+t_2}(r,s) = \int K_{t_1}(r,x)K_{t_2}(x,s)du(s).$$

This property is called the convolution property.

3. The kernel satisfies the following equations

$$\int K_t(r,s)du(s) = 1 \quad \text{and} \quad \int K_t(r,s)du(r) = 1,$$

that is, it is a probability density under measure du in either argument.

Our interest in diffusion kernels stems from the fact that these kernels satisfy the convolution property, allowing easy computation.

Definition 2. Let $\mathcal{B}^d = \{x \in \mathbb{R}^d : ||x|| < 1\}$ and let $\mathcal{S}^{d-1}$ represent the unit sphere, the boundary of $\mathcal{B}^d$. The d-dimensional Poisson kernel is defined as

$$P(r,s) = \frac{1-||r||^2}{||r-s||^2}, \quad s \in \mathcal{S}^{d-1}, \quad r \in \mathcal{B}^d.$$

The aforementioned expression is neither symmetric nor it has the desirable integration properties. Define now a kernel on $\mathcal{S}^{d-1} \times \mathcal{S}^{d-1}$ by

$$P_\rho(u,v) = P(\rho u, v) = \frac{1-\rho^2}{(1+\rho^2-2\rho(u\cdot v))^{d/2}},$$

$u, v \in \mathcal{S}^{d-1}$. This expression is symmetric and integrates to one with respect to the uniform measure on the d-dimensional unit sphere. The parameter ρ is called the *concentration parameter* of the kernel. This new expression defines the d-dimensional Poisson kernel.

When $d = 2$ the form of the Poisson kernel is

$$P(\rho u, v) = \frac{1-\rho^2}{1+\rho^2-2\rho u\cdot v}.$$

Since any vector v on the unit sphere corresponds uniquely to an angle $\theta \in [0, 2\pi)$ via the representation $v = (\cos\theta, \sin\theta)$, if u has corresponding angle ϕ, the Poisson kernel can be written as

$$P(\rho u, v) = \frac{1-\rho^2}{1+\rho^2-2\cos(\theta-\phi)}.$$

This expression is the wrapped Cauchy density. Golzy and Markatou [10] clarify the relationship of the Poisson kernel-based densities with other spherical distributions, see the aforementioned reference for further details on these connections. Here, it is worth mentioning that for a general dimension d the Poisson kernel-based density is the exit distribution on the sphere density obtained by [11], see Kato and Jones [11] for further details.

Mixtures of Poisson kernel-based densities were proposed for use in clustering and an important aspect of mixture models is that of their identifiability. Mixtures of Poisson kernel-based densities are identifiable in the following cases.

1. If the concentration parameter ρ is common among all components and known and the number of components in the mixture is unknown the mixture of Poisson kernel-based densities is identifiable.
2. If the number of components is known, the mixture of Poisson kernel-based densities is identifiable if the corresponding family of the densities $\{f(x, \rho_j, \mu_j) : 0 \leq \rho_j < 1, \mu_j \in \mathcal{S}^{d-1}\}$ is linearly independent.

4 Computation

The practical implementation of clustering algorithms requires attention of several aspects. One such aspect is deciding the number of clusters, while other aspects concern initialization, the stopping rule that is to be used, as well as the existence of software that implements the method. The literature includes several methods that can be used to estimate the number of clusters. Some of these methods are described in [2,9,17,19]. Anyone of these methods can be used. Graphical methods, for example the elbow plot, can also be used to estimate the number of clusters. The package `QuadratiK`, implemented in both, `R` and Python, can be used to identify data clusters using mixtures of Poisson kernel-based densities [18]. The package estimates the number of clusters using the elbow plot.

5 Conclusions

Mixtures of Poisson kernel-based densities are useful models for clustering data on the sphere, especially when the percentage of noise in the data is high or when there is a large overlap between different clusters. The functions incorporated in the recently developed `QuadratiK` package can be used to cluster spherical data using mixtures of Poisson kernel-based densities.

Acknowledgements. The author would like to thank Anran Liu, graduate student in the Department of Biostatistics, SPHHP, University at Buffalo, for the help she provided.

References

1. Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Clustering on the unit hypersphere using von mises-fisher distributions. J. Mach. Learn. Res. **6**, 1345–1382 (2005)
2. Biernacki, C., Celeux, G., Govaert, G.: Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 719–725 (2000)
3. Bijral, A.S., Breitenbach, M., Grudic, G.: Mixture of Watson distributions: A generative model for hyperspherical embeddings. In: Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 2, pp. 35–42 (2007)

4. Dey, S., Jana, N.: Inference on parameters of Watson distributions and application to classification of observations. J. Comput. Appl. Math. **403**, 113847 (2022)
5. Dhillon, I.S., Modha, D.S.: Concept decompositions for large sparse text data using clustering. Mach. Learn. **42**, 143–175 (2001)
6. Di Marzio, M., Fensore, S., Panzera, A., Taylor, C.C.: Kernel density classification for spherical data. Stat. Probab. Lett. **144**, 23–29 (2019)
7. Ding, Y., Markatou, M., Saraceno, G.: Poisson kernel-based tests for uniformity on the d-dimensional sphere. Stat. Sin. (in press) (2023)
8. Fernando Vera, J., Subiabre, R., Macías, R.: Clustering and geodesic scaling of dissimilarities on the spherical surface. J. Agric. Biol. Environ. Stat. (2024). https://doi.org/10.1007/s13253-023-00597-4
9. Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. **97**, 611–631 (2002)
10. Golzy, M., Markatou, M.: Poisson kernel-based clustering on the sphere: convergence properties, identifiability, and a method of sampling. J. Comput. Graph. Stat. **29**(4), 758–770 (2020)
11. Kato, S., Jones, M.: An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. Bernoulli **19**, 154–171 (2013)
12. Kent, J.T.: The Fisher-Bingham distribution on the sphere. J. R. Stat. Soc. Ser. B Stat. Method. **44**(1), 71–80 (1982)
13. Kim, S., SenGupta, A.: A three-parameter generalized von mises distribution. Stat. Pap. **54**, 685–693 (2013)
14. Kim, S., SenGupta, A.: Clustering methods for spherical data: An overview and a new generalization. In: Choi, D., et al. (Eds.) Proceedings of the Pacific Rim Statistical Conference for Production Engineering: Big Data, Production Engineering and Statistics. ICSA Book Series in Statistics, pp. 155–164. Springer, Singapore (2018)
15. Maitra, R., Ramler, I.P.: A k-mean-directions algorithm for fast clustering of data on the sphere. J. Comput. Graph. Stat. **19**(2), 377–396 (2010)
16. Mardia, K.V., Jupp, P.E.: Directional Statistics. Wiley Series in Probability and Statistics. Wiley, New York (2000)
17. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
18. Saraceno, G., Markatou, M., Mukhopadhyay, R., Golzy, M.: Goodness-of-fit and clustering of spherical data: the QuadratiK package in R and Python. arxiv2402.02290v1[stat.CO], https://doi.org/10.48550/arXiv.2402.02290 (2024)
19. Tibshirani, R., Walther, G.: Cluster validation by prediction strength. J. Comput. Graph. Stat. **14**, 511–528 (2005)
20. Zhong, S.: Efficient online spherical k-means clustering. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, vol. 5, pp. 3180–3185 (2005)

Aggregation of the Distortion Models Induced by the KL Divergence and Euclidean Distance

Ignacio Montes(✉)

Department of Statistics and O.R, University of Oviedo,
C/Federico García Lorca 18, 33007 Oviedo, Spain
imontes@uniovi.es

Abstract. Distortion or neighbourhood models are tools within the imprecise probability theory that allow to robustify a probability measure. These are built by considering the closed ball around a probability measure with a given radius and using a distorting function to compare probability measures. These include well-known models such as the linear vacuous, pari-mutuel or total variation models. In this contribution we focus on the distortion models that arise from considering the Euclidean distance or the Kullback-Leibler divergence as distorting functions, and analyse their behaviour under different aggregation rules: conjunction, disjunction or convex mixtures.

Keywords: Distortion models · Credal sets · Coherent lower previsions · Aggregation · Euclidean distance · Kullback-leibler divergence

1 Introduction

Robust statistics [5] aim to provide probabilistic tools that ensure the outcome of probabilistic or statistical procedures is not heavily influenced by the inputs. In this respect, a common approach to robustify a probability measure is by taking a neighbourhood around it. A well-known example is the ε-*contamination* model [5], later called *linear vacuous* model by Walley [13], based on a convex combination between the given probability (with weight $1-\varepsilon$) and any other probability (with weight ε).

More generally, a *distortion* or *neighbourhood* model [8,9] is constructed using the closed ball centred at the giving probability measure with a radius indicating the degree of robustness, and with respect to a function that compares probability measures. This approach includes not only the linear vacuous but some well-known models from the imprecise probability theory [1,13] such as the pari-mutuel model [7,11], total-variation model [4] and also vertical barrier models [2], which perform a nearly linear transformation of a probability measure.

A challenging problem when dealing with probabilistic models is the aggregation of information given by experts. In this respect, Walley [12] performs

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 286–293, 2024.
https://doi.org/10.1007/978-3-031-65993-5_35

a detailed analysis of the aggregation of probability measures and of sets of probability measures, proposing conjunction, disjunction or convex mixtures as aggregation rules, among others. The aggregation of linear vacuous, pari-mutuel or total variation models was explored in [3,7]. In this paper, we consider two distortion models recently introduced in [6], those arising from using as distorting function the Euclidean distance and the Kullback-Leibler (KL, for short) divergence, and analyse their behaviour under these three aggregation rules.

To achieve this goal, after formally introducing the main notions about distortion models in Sect. 2, we devote Sects. 3 and 4 to analyse the aggregation of the distortion models induced by the Euclidean distance and KL divergence, respectively. We conclude the paper with some final remarks in Sect. 5.

2 Distortion Models

We consider a finite possibility space $\mathcal{X} = \{x_1, \ldots, x_n\}$, and denote by $\mathcal{L}(\mathcal{X})$ the set of functions $f : \mathcal{X} \to \mathbb{R}$, called gambles, and by $\mathbb{P}^*(\mathcal{X})$ the set of probability measures on $\mathcal{P}(\mathcal{X})$ assigning strictly positive probability to non-empty events.

Given a probability measure P_0, a distortion parameter δ and a function d comparing probability measures, a *distortion model* [8] is defined as $B_d^\delta(P_0) = \{P \mid d(P, P_0) \leq \delta\}$. When d is continuous and convex, $B_d^\delta(P_0)$ is a closed and convex set of probability measures, called *credal set*. Under this assumption, its lower envelope, called *coherent lower prevision* and given, for any $f \in \mathcal{L}(\mathcal{X})$, by

$$\underline{P}_d(f) = \min\{E_P(f) \mid d(P, P_0) \leq \delta\} = \min\left\{E_P(f) \mid P \in B_d^\delta(P_0)\right\},$$

allows to express the distortion model as $B_d^\delta(P_0) = \left\{P \mid E_P(f) \geq \underline{P}_d(f)\ \forall f \in \mathcal{L}(\mathcal{X})\right\}$. Its associated *coherent upper prevision* is defined by $\overline{P}_d(f) = -\underline{P}_d(-f)$.

There are a number of aggregation rules that can be used to aggregate closed and convex set of probability measures [12]. Given two distortion models $B_d^{\delta_1}(P_0^1)$ and $B_d^{\delta_2}(P_0^2)$, we consider the following rules:

Conjunction It is given by $B_d^{\delta_1}(P_0^1) \cap B_d^{\delta_2}(P_0^2)$. Whenever this intersection is non-empty, its lower envelope $\underline{P}$ satisfies $\underline{P} \geq \max\{\underline{P}_d^1, \underline{P}_d^2\}$, but the inequality can be strict.

Disjunction It is given by $B_d^{\delta_1}(P_0^1) \cup B_d^{\delta_2}(P_0^2)$. The lower envelope of the union, denoted by $\underline{P}$, satisfies $\underline{P} = \min\left\{\underline{P}_d^1, \underline{P}_d^2\right\}$.

Convex mixtures For a fixed $\varepsilon \in (0,1)$, it is given by $\varepsilon B_d^{\delta_2}(P_0^1) + (1-\varepsilon)B_d^{\delta_2}(P_0^2)$.

For each of them, we say that the distortion model is closed under the aggregation rule if the resulting set of probability measures can be represented as a distortion model of the same family: $B_d^{\delta^*}(P_0^*)$ for some δ^* and P_0^*. Generally, the disjunction produces a non-convex set, prompting us to inquire whether there is a minimal $B_d^{\delta^*}(P_0^*)$ including the disjunction, called *undominated outer approximation* [10].

It is not difficult to show that, under fairly general conditions, the disjunction is not preserved by the distortion model.

Proposition 1. *If d is continuous, closed, it satisfies $B_d^{\delta_1}(P_0^1) = B_d^{\delta_2}(P_0^2)$ if and only if $P_0^1 = P_0^2$, and $\delta_1 = \delta_2$ and $B_d^{\delta_1}(P_0) \subset B_d^{\delta_2}(P_0)$ for $\delta_1 < \delta_2$, then the distortion model is not preserved under disjunction in general.*

Next, we consider the distortion models induced by the Euclidean distance and KL divergence, and we analyse if they are closed under the proposed rules.

3 Aggregation of Euclidean Models

Giving two probability measures P and Q, their Euclidean distance is given by:

$$d_E(P,Q) = \sqrt{\sum_{i=1}^{n} \big(P(\{x_i\}) - Q(\{x_i\})\big)^2}.$$

It induces the distortion model $B_E^\delta(P_0)$ [6] and gives rise to the coherent lower prevision $\underline{P}_E(f) = E_P(f) - \delta\sqrt{n}S_f$ for any $f \in \mathcal{L}(\mathcal{X})$, where S_f is given by $S_f^2 = \frac{1}{n}\sum_{i=1}^n \big(f(x_i) - \overline{f}\big)^2$ and $\overline{f} = \frac{1}{n}\sum_{i=1}^n f(x_i)$. Since d_E is continuous and convex, $B_E^\delta(P_0) = \{P \mid E_P(f) \geq \underline{P}_E(f)\ \forall f \in \mathcal{L}(\mathcal{X})\}$ and it is characterised by its extreme points, given by $\{P \mid d_E(P,P_0) = \delta\}$. Next, we investigate the behaviour of the Euclidean model under different aggregation procedures.

Conjunction. First of small, it is not difficult to show that the Euclidean model is not closed under conjunction since the intersection of spheres is, in general, no longer a sphere. We next sketch an example.

Example 1. Consider $\mathcal{X} = \{x_1, x_2, x_3\}$, the probability measures $P_0^1 = (0.5, 0.3, 0.2)$ and $P_0^2 = (0.5, 0.2, 0.3)$, the distortion parameters $\delta = \delta_1 = \delta_2 = \sqrt{2}/10$ and the Euclidean models $B_E^\delta(P_0^1)$ and $B_E^\delta(P_0^2)$ they induce. A graphical representation of both models, as well as their conjunction, can be seen in Fig. 1. Clearly, the intersection is no longer an sphere. Even more, in the intersection the difference $\overline{P}(\{x\}) - \underline{P}(\{x\})$ is grater for x_1 than for x_2, while in an Euclidean model it should be constant for any element. ♦

Disjunction. Since the Euclidean distance satisfies the properties of Prop. 1, the Euclidean model is not preserved by disjunction. Next we show an example where the disjunction of two Euclidean model is not convex.

Example 2. Let us continue with Example 1. Consider the probability measures

$$P_1 = \left(0.5 + \delta\sqrt{2/3}, 0.3 - \delta/2\sqrt{2/3}, 0.2 - \delta/2\sqrt{2/3}\right) \in B_E^\delta(P_0^1).$$
$$P_2 = \left(0.5 + \delta\sqrt{2/3}, 0.2 - \delta/2\sqrt{2/3}, 0.3 - \delta/2\sqrt{2/3}\right) \in B_E^\delta(P_0^2).$$

P_1 and P_2 satisfy $d_E(P_1, P_0^1) = d_E(P_2, P_0^2) = \delta$, $\overline{P}_E^1(\{x_1\}) = P_1(\{x_1\})$ and $\overline{P}_E^2(\{x_1\}) = P_2(\{x_1\})$. However, $P = \frac{1}{2}P_1 + \frac{1}{2}P_2$ satisfies $d_E(P, P_0^1) = d_E(P, P_0^2) = \frac{\sqrt{2.5}}{10} > \frac{\sqrt{2}}{10} = \delta$. Hence, $P \notin B_E^\delta(P_0^1) \cup B_E^\delta(P_0^2)$.

A graphical representation can be seen in Fig. 2 (left-hand side). ♦

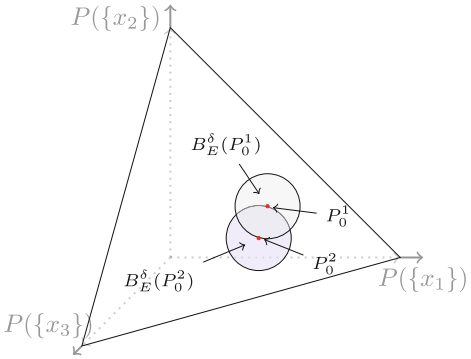


Fig. 1. Graphical representation of $B_E^{\delta}(P_0^1)$, $B_E^{\delta}(P_0^2)$ and their conjunction

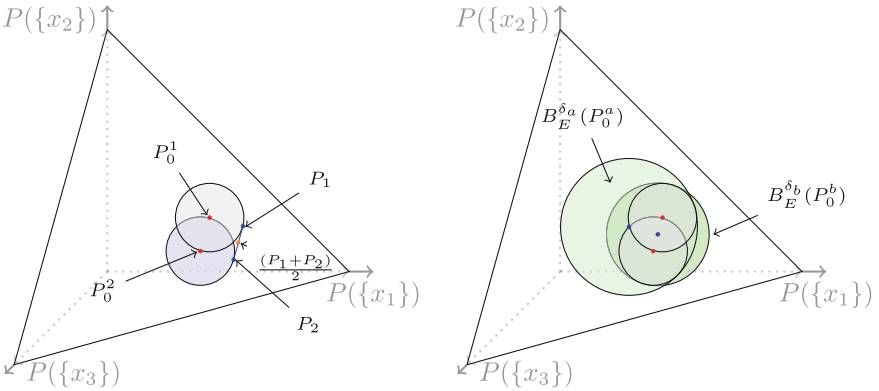


Fig. 2. Graphical representation of $B_E^{\delta}(P_0^1)$ and $B_E^{\delta}(P_0^2)$ showing that their disjunction is not convex (left) and two different undominated outer approximations $B_E^{\delta_a}(P_0^a)$ and $B_E^{\delta_b}(P_0^b)$ (right)

The right-hand side of Fig. 2 also shows that there is not a unique undominated outer approximation of the disjunction using an Euclidean model. The next result shows that there is a unique Euclidean model outer approximating the disjunction with a minimal distortion parameter.

Proposition 2. *Let* $B_E^{\delta_1}(P_0^1), B_E^{\delta_2}(P_0^2) \subseteq \mathbb{P}^*(\mathcal{X})$ *be two Euclidean models and*

$$\delta^* = \min\left\{\delta \mid B_E^{\delta_1}(P_0^1) \cup B_E^{\delta_2}(P_0^2) \subseteq B_E^{\delta^*}(P) \text{ for some } P \in \mathbb{P}^*(\mathcal{X})\right\}.$$

Whenever $\delta^* < \sqrt{\frac{n}{n-1}} \min_{x\in\mathcal{X}} \left\{P_0^1(\{x\}), P_0^2(\{x\})\right\}$, *there is a unique* $P^* \in \mathbb{P}^*(\mathcal{X})$ *such that* $B_E^{\delta_1}(P_0^1) \cup B_E^{\delta_2}(P_0^2) \subseteq B_E^{\delta^*}(P^*)$.

In Fig. 2, the ball $B_E^{\delta^*}(P^*)$ obtained in the previous result was denoted by $B_E^{\delta_b}(P_0^b)$.

Mixtures. Finally we prove that the Euclidean model is closed under convex mixtures.

Proposition 3. *Consider* $P_0^1, P_0^2 \in \mathbb{P}^*(\mathcal{X})$ *and* $\delta > 0$ *such that* $B_E^{\delta_1}(P_0^1) \subseteq \mathbb{P}^*(\mathcal{X})$ *and* $B_E^{\delta_2}(P_0^2) \subseteq \mathbb{P}^*(\mathcal{X})$. *Given* $\varepsilon \in (0,1)$, *it holds that* $\varepsilon B_E^{\delta_1}(P_0^1) + (1-\varepsilon)B_E^{\delta_2}(P_0^2) = B_E^{\delta_\varepsilon}(P_0^\varepsilon)$, *where* $\delta_\varepsilon = \varepsilon\delta_1 + (1-\varepsilon)\delta_2$ *and* $P_0^\varepsilon = \varepsilon P_0^1 + (1-\varepsilon)P_0^2$.

4 Aggregation of KL Models

Given two probability measures P and Q, their KL divergence is given by:

$$D_{KL}(P,Q) = \sum_{i=1}^{n} P(\{x_i\}) \log\left({}^{P(\{x_i\})}/_{Q(\{x_i\})}\right).$$

The KL divergence is continuous and convex, hence the distortion model it induces, denoted as $B_{KL}^{\delta}(P_0)$ and called KL-model [6], can be expressed as $\{P \mid E_P(f) \geq \underline{P}_{KL}(f)\ \forall f \in \mathcal{L}(\mathcal{X})\}$, where $\underline{P}_{KL}$ is the lower envelope of the ball.

As shown in [6], computing $\underline{P}_{KL}(f)$ analytically is intricate and should be done by solving an optimisation problem. When $f = I_A$ for some event A, $\underline{P}(A) = \alpha P_0(A)$, where α is the unique solution in the interval $[0,1]$ of the equality:

$$\alpha P_0(A)\log(\alpha) + (1-\alpha P_0(A))\log\left(\frac{1-\alpha P_0(A)}{1-P_0(A)}\right) = \delta. \tag{1}$$

Moreover, the set of extreme points of $B_{KL}^{\delta}(P_0)$ is $\{P \mid D_{KL}(P,P_0) = \delta\}$.

We now investigate the behaviour of the KL-model under aggregation.

Conjunction. We start showing that the KL-model is not closed under conjunction.

Example 3. Consider $\mathcal{X} = \{x_1,x_2,x_3\}$, the probability measures $P_0^1 = (0.4,0.3,0.3)$ and $P_0^2 = (0.3,0.4,0.3)$, and the distortion parameters $\delta = \delta^1 = \delta^2 = 0.1$. The balls $B_{KL}^{\delta}(P_0^1)$ and $B_{KL}^{\delta}(P_0^2)$ are graphically depicted in Fig. 3.

Let us denote by $\underline{P}^{\cap}$ the lower prevision associated with $B_{KL}^{\delta}(P_0^1) \cap B_{KL}^{\delta}(P_0^2)$, which satisfies $\underline{P}^{\cap} \geq \max\{\underline{P}_{KL}^1, \underline{P}_{KL}^2\}$.

Assume that $\underline{P}^{\cap}$ is associated with a KL-model induced by $P_0^{\cap}$ and $\delta^{\cap}$. The symmetry of the balls implies $\underline{P}^{\cap}(\{x_1\}) = \underline{P}^{\cap}(\{x_2\})$, meaning that $P_0^{\cap}(\{x_1\}) = P_0^{\cap}(\{x_2\})$. Hence, $P_0^{\cap}$ would be given by $P_0^{\cap} = (p,p,1-2p)$ for some $p \in [0,1]$. It can be seen that $\underline{P}^{\cap}(\{x_3\}) = 0.1204$ and $\overline{P}^{\cap}(\{x_3\}) = 0.5089$, and these values are attained only in the probability measures $Q_1 = (0.43978, 0.43978, 0.1204)$ and $Q_2 = (0.2455, 0.2455, 0.5089)$, respectively. Solving the equalities $D_{KL}(Q_1, P_0^{\cap}) = D_{KL}(Q_2, P_0^{\cap}) = \delta^{\cap}$ we obtain $p = 0.3489$. Hence $P_0^{\cap} = (0.3489, 0.3489, 0.30217)$ and $\delta^{\cap} = 0.0927$. However, even though $Q = (0.1915, 0.2995, 0.5089) \in B_{KL}^{\delta}(P_0^1) \cap B_{KL}^{\delta}(P_0^2)$ and $\underline{P}_{KL}(\{x_1\}) = Q(\{x_1\})$, it holds that $D_{KL}(Q, P_0^{\cap}) = 0.1047 > \delta^{\cap}$, a contradiction. We conclude that $\underline{P}^{\cap}$ is not associated with a KL-model. ♦

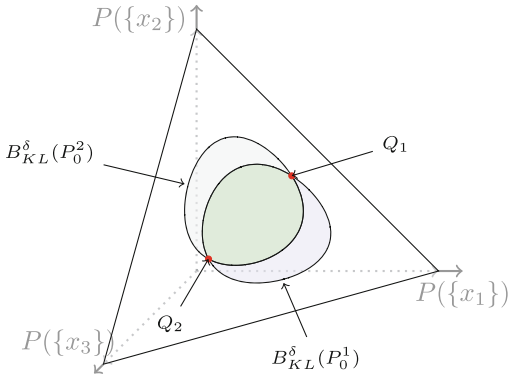


Fig. 3. Graphical representation of $B^{\delta}_{KL}(P^1_0)$ and $B^{\delta}_{KL}(P^2_0)$ in Example 3, as well as their conjunction, in green

Disjunction. Since the KL-divergence satisfies the properties in Prop. 1, the KL-model is not closed under disjunction. Indeed, the same example illustrates that the disjunction of two KL-models may not be convex. Next we show that there is neither a unique outer approximation of the disjunction using a KL-model.

Example 4. Let us continue with Example 3. On the one hand, the minimum distortion parameter δ^* such that there exists a probability measure $P^* \in \mathbb{P}^*(\mathcal{X})$ satisfying $B^{\delta^*}_{KL}(P^*_0) \supseteq B^{\delta}_{LK}(P^1_0) \cup B^{\delta}_{LK}(P^2_0)$ is $\delta^* = 0.15381$ and the probability measure must be $P^*_0 = (0.3552, 0.3552, 0.2896)$. The ball $B^{\delta^*}_{KL}(P^*_0)$ has been depicted in green in the left part of Fig. 4.

On the other hand, we consider $P^{'}_0 = (1/3, 1/3, 1/3)$, and we look for the minimum distortion parameter δ' such that the lower prevision $\underline{P}^{'}_{KL}$ induced by $B^{\delta'}_{KL}(P^{'}_0)$ satisfies

$$\underline{P}^{'}_{KL}(\{x_1\}) = \min\left\{\underline{P}^1_{KL}(\{x_1\}), \underline{P}^2_{KL}(\{x_1\})\right\},$$
$$\overline{P}^{'}_{KL}(\{x_1\}) = \max\left\{\overline{P}^1_{KL}(\{x_1\}), \overline{P}^2_{KL}(\{x_1\})\right\},$$

getting $\delta' = 0.173594$. The ball $B^{\delta^{'}}_{KL}(P^{'}_0)$ is depicted in yellow in the right part of Fig. 4. It follows that none of $B^{\delta^*}_{KL}(P^*_0)$ and $B^{\delta^{'}}_{KL}(P^{'}_0)$ is included in the other. Since we cannot decrease the distortion parameters δ^* and $\delta^{'}$, these are two different undominated outer approximations of the disjunction. ♦

Mixtures. Finally, the next example shows that the KL-model is also not closed under a convex combinations.

Example 5. Consider the possibility space $\mathcal{X} = \{x_1, x_2, x_3\}$, the probability measures $P^1_0 = (0.4, 0.3, 0.3)$, $P^2_0 = (0.3, 0.4, 0.3)$ and $P^3_0 = (0.3, 0.3, 0.4)$, the distortion parameter $\delta = 0.1$ and the KL-models $B^{\delta}_{KL}(P^1_0)$, $B^{\delta}_{KL}(P^2_0)$ and

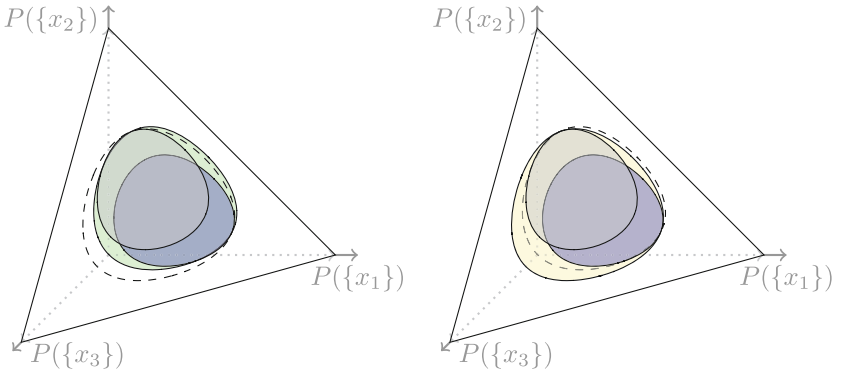


Fig. 4. Graphical representation of $B^{\delta}_{KL}(P^1_0)$ and $B^{\delta}_{KL}(P^2_0)$ in Example 4 together with two different undominated outer approximations

$B^{\delta}_{KL}(P^3_0)$. If their average were a KL-model $B^{\delta^*}_{KL}(P_0)$, P_0 should be uniform because its induced lower probability $\underline{P}$ satisfies:

$$\underline{P}(\{x_1\}) = \underline{P}(\{x_2\})) = \underline{P}(\{x_3\}) = 0.112736.$$
$$\underline{P}(\{x_1, x_2\}) = \underline{P}(\{x_1, x_3\}) = \underline{P}(\{x_2, x_3\}) = 0.44956.$$

In particular, Eq. (1) becomes

$$0.1390195 = \frac{2}{3}0.112736 + \frac{1}{3}0.19158 = \underline{P}(\{x_1\}) = \alpha P_0(\{x_1\}) = \frac{\alpha}{3},$$

meaning that $\alpha = 0.4170585$. From the proof of [6, Thm.7] we deduce that $\underline{P}(\{x_1\})$ is attained in the probability measure

$$P = \left(\frac{\alpha}{3}, \frac{\beta}{3}, \frac{\beta}{3}\right) = (0.1390195, 0.43049025, 0.43049025),$$

where $\beta = \frac{1-\alpha/3}{1-1/3} = 1.291471$. We conclude that $\delta^* = D_{KL}(P, P_0) = 0.0986465$. However, with this distortion parameter we obtain that:

$$\min\left\{P(\{x_1, x_2\}) \mid D_{KL}(P, P_0) = \delta^*\right\} = 0.4499001,$$

but this value is different than 0.44956, hence we conclude that the average of $B^{\delta}_{KL}(P^1_0)$, $B^{\delta}_{KL}(P^2_0)$ and $B^{\delta}_{KL}(P^3_0)$ cannot be a KL-model. ♦

5 Conclusions

In this contribution we have studied the aggregation of Euclidean and KL distortion models using the conjunction, disjunction and convex mixtures. The results reveal that their behaviour under the aggregation rules is questionable: unlike

the linear vacuous and pari-mutuel models [3,7], the Euclidean and KL-models are not closed under conjunction. Moreover, they are also not closed under disjunction and they do not possess a unique undominated outer approximation, contrary to the linear vacuous and pari mutuel model [3]. Lastly, even though the Euclidean, linear vacuous, pari mutuel and total variation models are closed under convex mixtures, the same cannot be said for the KL-model.

For future research, we must analyse in the future how these two models behave when combining different models into a joint one in a multivariate setting to complete the study given in [3]. Moreover, even though these two models do not have a good behaviour under conjunction, disjunction or convex mixtures, we may analyse their aggregation with the other rules proposed in [12].

Acknowledgements. This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "FEDER/UE". I also acknowledge the careful reading and the many interesting suggestions posed by the reviewers.

References

1. Augustin, T., Coolen, F., De Cooman, G., Troffaes, M. (eds.): Introduction to Imprecise Probabilities. Wiley Series in Probability and Statistics. Wiley
2. Corsato, C., Pelessoni, R., Vicig, P.: Nearly-Linear uncertainty measures. Int. J. Approx. Reason. **114**, 1–28 (2019)
3. Destercke, S., Montes, I., Miranda, E.: Processing distortion models: a comparative analysis. Int. J. Approx. Reason. **145C**, 91–120 (2022)
4. Herron, T., Seidenfeld, T., Wasserman, L.: Dividive conditioning: further results on dilation. Philos. Sci. **66**, 411–444 (1997)
5. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
6. Montes, I.: Neighbourhood models induced by the Euclidean distance and the Kullback-Leibler divergence. In: Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications. Proceedings of Machine Learning Research, vol. 215, pp. 367–378 (2023)
7. Montes, I., Miranda, E., Destercke, S.: Pari-mutuel probabilities as an uncertainty model. Inf. Sci. **481**, 550–573 (2019)
8. Montes, I., Miranda, E., Destercke, S.: Unifying neighbourhood and distortion models: Part I-New results on old models. Int. J. Gen Syst **49**(6), 602–635 (2020)
9. Montes, I., Miranda, E., Destercke, S.: Unifying neighbourhood and distortion models: Part II-New models and synthesis. Int. J. Gen Syst **49**(6), 636–674 (2020)
10. Montes, I., Miranda, E., Vicig, P.: 2-monotone outer approximations of coherent lower probabilities. Int. J. Approx. Reason. **101**, 181–2015 (2018)
11. Pelessoni, R., Vicig, P., Zaffalon, M.: Inference and risk with the pari-mutuel model. Int. J. Approx. Reason. **51**, 1145–1158 (2010)
12. Walley, P.: The Elicitation and Aggregation of Beliefs. Warwick Research Report 23 (1982)
13. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Regularised Optimal Self-transport Is Approximate Gaussian Mixture Maximum Likelihood

Gilles Mordant(✉)

IMS,Universität Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany
gilles.mordant@uni-goettingen.de

Abstract. We investigate the link between regularised self-transport problems and maximum likelihood estimation in Gaussian mixture models (GMM). This link suggests that self-transport followed by a clustering technique leads to principled estimators at a reasonable computational cost. Also, robustness, sparsity and stability properties of the optimal transport plan arguably make the regularised self-transport a statistical tool of choice for the GMM.

Keywords: Optimal transport · Clustering · Gaussian mixture model

1 Introduction

Clustering is a standard problem in statistics and data science. Given a set of observations, the aim is to separate these observations in K groups that are similar. A typical model for which clustering is required is that of Gaussian mixtures; its definition is recalled in Sect. 1.1. Even though this model has a long history, answering the natural questions pertaining to the estimation of its parameters is still an active research endeavour, see [1,5,8] and the references therein for a more complete overview.

In a recent line of work [2,3,6] investigated the use of entropy-regularised self-transport to perform manifold learning; see Sect. 1.2 for an introduction to regularised optimal transport. In a subsequent paper, [9] considered the impact of using quadratically regularised optimal transport (QOT) for the same purposes, stressing the importance of the sparsity of the obtained transport plan for local adaptivity. In that last work, the authors exhibited the usefulness of applying spectral clustering on the optimal transport plan to recover the clusters of Gaussian Mixtures Models. The utility of bistochastic projection was also independently observed by [4]. Therein, however, the projection onto hollow bistochastic matrices is not considered, even though it turns out to be an important feature. Also, the latter reference does not consider the problem from a statistical viewpoint.

In this note, we exhibit the reason why a method based on QOT performs well in practice by relating it to maximum likelihood estimation of GMM. This is done in Sect. 2. Before doing so, we first provide preliminary background. Section 3 contains the proofs and Sect. 4 concludes.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 294–301, 2024.
https://doi.org/10.1007/978-3-031-65993-5_36

1.1 Gaussian Mixtures

Let us denote by Δ_K the K-simplex.[1] We recall that d-dimensional Gaussian mixtures model with K components can be thought of as the result of two-stage process. First, one selects the cluster or subgroup at random, which is modelled as sampling from the multinomial $\mathrm{Mult}(1;\theta_1,\dots,\theta_K)$, where $\theta \in \Delta_K$ is the vector encoding the group proportions. Then, conditionally on the selected cluster, one draws from the Gaussian random variable with corresponding parameters. We denote the mean vectors by μ_k and the covariance matrices by Σ_k, $k \in \{1,\dots K\}$. For now, let us further set $\Sigma_k = \sigma^2 I_d$, for each k and assume that σ^2 is fixed and known. The cluster membership of the observation i is encoded in the random vector Z_i of size K containing only zeros except at the position of the cluster where it is one.

The observations thus are pairs $(Z_i, Y_i), 1 \le i \le n$ and maximising the *full* log-likelihood is maximising

$$\ell_n(Z;K;\mu_1,\dots,\mu_K;\theta) := \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik}\left(\log\theta_k - \frac{1}{2\sigma^2}\|Y_i-\mu_k\|^2\right) - nd\log\sigma. \quad (1)$$

Solving this maximisation problem is challenging, in particular because the Z_i's are not observed. We now recall some basics of discrete, regularised optimal transport.

1.2 Regularised Optimal Transport

In (discrete) optimal transport, one is given two sets of points $\mathbb{X} := \{x_1,\dots,x_n\}$ and $\mathbb{Y} := \{y_1,\dots,y_n\}$ and the corresponding empirical measures. Given a cost function $c : \mathbb{X}\times\mathbb{Y} \to \mathbb{R}$, there is a corresponding $n \times n$ cost matrix C. The regularised optimal transport problem then reads,

$$\min_{\pi\in B_n} \langle \pi, C\rangle_F + \varepsilon\Phi(\pi),$$

where B_n is the set of bistochastic matrices, ε is a regularisation parameter and Φ is a convex function, usually a f-divergence with respect to the uniform measure on $\mathbb{X}\times\mathbb{Y}$. Let $\langle\cdot,\cdot\rangle_F$ and $\|\cdot\|_F$ denote the Frobenius inner product and norm. In this note, we consider the two most classical functions Φ.

- First, $\Phi(\pi) = \sum_{i,j}\pi_{i,j}\log(\pi_{i,j})$, with the convention $0\log 0 = 0$. The problem in (2) for that particular choice is then called the entropy-regularised optimal transport.
- Second, the choice $\Phi(\pi) = \sum_{i,j}\pi_{i,j}^2/2$ gives birth to the quadratically regularised optimal transport problem.

A classical choice for the cost function, provided that the points belong to $\mathbb{R}^d$, is the squared Euclidean distance.

[1] The K-simplex is defined as $\Delta_k := \left\{x \in \mathbb{R}_+^K : \sum_{k=1}^K x_k = 1\right\}$.

Finally, let us stress an important property of entropy or quadratically regularised optimal transport, when B_n is replaced by the set of bistochastic matrices with zero diagonal, call the latter set B_n°. In that case, the problem exhibits robustness to *heteroskedastic* noise in high dimensions. For more clarity, consider a set of points x_i's in high dimension that is corrupted with noise, i.e., one observes

$$\tilde{x}_i = x_i + \eta_i,$$

where $\mathbb{E}\eta_i = 0$ and $\mathbb{E}\eta_i\eta_i^\top = \Sigma_i$, where the $\Sigma_i's$ can be different for all indices. In the sequel, the noise random variables are assumed independent and independent of everything else. In high dimensional settings, for $i \neq j$ and well-behaved[2] Σ_i's, it holds that

$$\|\tilde{x}_i - \tilde{x}_j\|^2 \approx \|\eta_i\|^2 + \|\eta_j\|^2 + \|x_i - x_j\|^2.$$

Therefore, if one enforces the transport plan to be zero on the diagonal, the additional terms arising from the heteroskedastic noise is a rank-one perturbation, which is absorbed by the optimal potentials. A precise version of this result adapted to the object of this paper is stated in Sect. 3. Therefore, the self-transport problem that we will be considering is

$$\pi^*_{\varepsilon,\Phi} = \operatorname*{arg\,min}_{\pi \in B_n^\circ} \langle \pi, C\rangle_F + \varepsilon\Phi(\pi). \tag{2}$$

Before moving to the next section, we stress the fact that quadratically optimised transport is equivalent [9] to a metric projection problem for $\varepsilon > 0$, i.e.,

$$\pi^*_{\varepsilon,Q}(C) := \operatorname*{arg\,min}_{\pi \in B_n^\circ} \langle C, \pi\rangle_F + \frac{\varepsilon}{2}\|\pi\|_F^2 = \operatorname*{arg\,min}_{\pi \in B_n^\circ} \left\|\pi - \frac{-C}{\varepsilon}\right\|_F^2,$$

which is easily seen by expanding the square. It further is a classical result that $C \mapsto \pi^*_{\varepsilon,Q}(C)$ is a Lipschitz function of C.

2 Main Results

We now turn to the main contributions of this note. In the sequel, $H(\theta)$ denotes the entropy of the non-negative vector θ and $\|A\|_2$ denotes the largest singular value of the matrix A.

2.1 Isotropic Gaussian Mixtures

Theorem 1. *Consider a sample of an isotropic Gaussian mixture model* $(Y_i)_{i=1}^n$ *with known variance* σ^2 *as above. Set*

$$C_{ij} := \frac{1}{2}\|Y_i - Y_j\|^2.$$

[2] See [3] or [9] for a precise statement of the conditions. In essence, they mean that the noise should be spherical enough and not concentrate in certain directions.

and assume that θ and K are known. Then, the problem of maximizing (1) *is asymptotically equivalent to*

$$\max_{Z\in\{0,1\}^{n\times K}} \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik}\ln\theta_k - \sum_{i=1}^{n} Z_{ik}\left(\|Y_i-\mu_k\|^2 + O_p\left(\frac{\sqrt{d}}{\sqrt{n}}\right)\right) - nd(\log\sigma-\sigma^2), \tag{3}$$

for the choice $\varepsilon = n\sigma^2\, H(\theta)/K$.

The proof is deferred to Sect. 3.

Remark 1. As announced in the title, this theorem means that quadratically regularised (hollow) self-transport approximates maximum likelihood estimation in a Gaussian mixture model for a proper choice of the regularisation parameter ε.

Naturally, the vector θ and K are unknown, but the order of magnitude of ε might be estimated. For the parameter K, it is important to note that the optimal solution π^* of (2) is such that

$$\pi^*_{ij} \approx \sum_{k=1}^{K} \mathbb{1}_{i\neq j} Z_{ik} Z_{jk} (N_k - 1)^{-1},$$

where the approximate equality sign is a consequence of the error terms in (3). The number of clusters will thus reflect in the rank of π^*. Altogether, a clustering method applied to π^* is thus a natural way to recover the different components of the mixture. From these clusters, one can then estimate K, θ and σ and verify that these values are in line with the optimal parameter ϵ.

2.2 Robustness to Heteroskedasticity

The results of Sect. 2.1 hold under the assumption that the Gaussians at play are isotropic with variance parameter σ^2. As already hinted at in Sect. 1.2, quadratically regularised hollow self-transport behaves well under heteroskedastic noise. This enables to extend the usefulness of the approach to cases of the different Σ_k's, provided that there exists $\sigma^2 > 0$ and a decomposition

$$\Sigma_k = \sigma^2 I_d + \tilde{\Sigma}_k, \qquad k \in \{1,\dots,K\}, \tag{4}$$

with each $\tilde{\Sigma}_k$ being positive semi-definite. This condition is violated if at least one Σ_k is merely positive *semi*-definite, it is thus not very restrictive. Under this decomposition, we can see the observations $(Y_i)_{i=1}^n$ as given by

$$Y_i = \tilde{Y}_i + \eta_i,$$

where the Y_i's are observations of an isotropic GMM and the noise vectors η_i are such that

$$\eta_i | Z_{i,k} \sim \mathcal{N}\left(0, \sum_{k=1}^{K} Z_{i,k}\Sigma_k\right).$$

We then define C and $\tilde{C}$ to be the pairwise squared Euclidean matrices of the Y_i's and the $\tilde{Y}_i$'s, respectively. With this notation, we can state the following theorem.

Theorem 2. *Take the heteroskedastic GMM as introduced above. Assume that there exist a constant* C_μ *such that* $\max_{1\leq k,k'\leq K} \|\mu_k - \mu_{k'}\| \leq C_\mu$. *Further assume that* $\|\tilde{\Sigma}_k\|_2 \leq C_\Sigma, \forall k \in \{1,\dots,K\}$. *Then,*

$$\left\|\pi^*_{\varepsilon,Q}(C) - \pi^*_{\varepsilon,Q}(C)\right\|_F \leq \frac{1}{\varepsilon}\left\|E\right\|_F,$$

where E *is a centred error matrix for which each entry is* $\mathrm{O}_p\left(d^{1/2}\right)$.

Remark 2. Remark that the order of the entries of the cost matrix is $\mathrm{O}_p(d)$, so that the errors indeed have a limited impact. We further stress that the right-hand side in the inequality above is an upper bound; it is difficult to assess whether it is sharp. When proving Prop. 3, one easily sees that more assumptions on the eigendecomposition of the $\tilde{\Sigma}_k$'s coupled with conditions on the direction of the vectors $\mu_k - \mu_{k'}$, $1 \leq k, k' \leq K$, could guarantee better bounds.

Proof. Proposition 3 below states that $C - \tilde{C}$ is the sum of a rank-one matrix, a diagonal matrix and some stochastic matrix that is $\mathrm{O}_p\left(d^{1/2}\right)$. Remark that a rank one-one perturbation and a diagonal perturbation of the cost matrix leaves the optimiser of (2) invariant. Thus, the difference between the solutions for C and $\tilde{C}$ only depends on the $\mathrm{O}_p\left(d^{1/2}\right)$ matrix. The Lipschitzianity of the metric projection recalled in Sect. 1.2 yields the claim. □

3 Proofs

3.1 Proof of Theorem 1

An estimator of (1) that removes the problem of estimating the μ_k's —even though still intractable as the Z_i's are unknown —is

$$\tilde{\ell}_n(Z;K;\theta) := \sum_{i=1}^n \sum_{k=1}^K Z_{ik} \ln \theta_k - \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \frac{Z_{ik}Z_{jk}}{N_k - 1} \mathbb{1}_{i\neq j} \frac{1}{2\sigma^2} \|Y_i - Y_j\|^2 - nd(\log\sigma - \sigma^2), \tag{5}$$

where N_k is the unobserved number of elements in the kth cluster. We first establish that this estimator is actually fulfilling its purpose.

Proposition 1. *It holds that*

$$\|Y_i - \mu_k\|^2 = \sum_{j=1}^n \frac{Z_{jk}}{N_k - 1} \mathbb{1}_{i\neq j} \frac{1}{2\sigma^2} \|Y_i - Y_j\|^2 - d\sigma^2 + O_p\left(\frac{\sqrt{d}}{\sqrt{n}}\right).$$

For comparison of the errors magnitude, notice that $\|Y_i - \mu_k\|^2 = O_p(d)$.

Proof. An expansion of squares yields

$$\|Y_i - \mu_k\|^2 = \|Y_i - Y_j\|^2 - 2\langle Y_i - \mu_k, \mu_k - Y_j\rangle - \|\mu_k - Y_j\|^2,$$

Note that, for $j = i$, the error is $\|\mu_k - Y_i\|^2$, which has a different sign from that of $-\|\mu_k - Y_j\|^2$ for the other cases. This explains the presence of the indicator variable. Then, basic computations yield

$$\sum_{\substack{j=1\\ j\neq 1}}^{n} \frac{Z_{jk}}{N_k - 1}\langle Y_i - \mu_k, \mu_k - Y_j\rangle \Big| (Y_i, Z_1, \ldots, Z_n) \sim \frac{1}{\sqrt{N_k - 1}}\mathcal{N}\Big(0, \sigma^2 \sum_{\ell=1}^{d}(Y_{i,\ell} - \mu_{k,\ell})^2\Big)$$

$$\text{as well as } \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{Z_{jk}}{N_k - 1}\|\mu_k - Y_j\|^2 \Big| (Z_1, \ldots, Z_n) \sim \frac{\sigma^2}{N_k - 1}\chi^2_{d\times(N_k-1)}.$$

□

The claim follows from a straightforward use of the laws of total expectation and variance.

Proposition 1 formalises the intuition that maximising the alternative (5) instead of (1) comes at little cost but removes the problem of estimating the μ_k's. To simplify notation, let us now set $\pi_{ij} := \sum_{k=1}^{K} \mathbb{1}_{i\neq j} Z_{ik} Z_{jk} (N_k - 1)^{-1}$.

Proposition 2. *The matrix π is a symmetric bistochastic matrix with all diagonal elements equal to zero.*

Proof. One easy sees that π is a (sparse) matrix with positive entries. The indicator in the definition ensures that the matrix is hollow. Further, for each $i \in \{1, \ldots, n\}$,

$$\sum_{j=1}^{n}\sum_{k=1}^{K} \mathbb{1}_{i\neq j} Z_{ik} Z_{jk} (N_k - 1)^{-1} = \sum_{k=1}^{K} (N_k - 1)^{-1} Z_{ik} \sum_{j=1}^{n} \mathbb{1}_{i\neq j} Z_{jk} = 1.$$

□

With the notation set before, Eq. (5) can thus be rewritten and obtain that $\sum_{k=1}^{K}\sum_{i=1}^{n} Z_{ik} \ln\theta_k = \sum_{k=1}^{K} N_k \ln\theta_k$. In view of Eq. (2), thus remains to relate $\sum_{k=1}^{K} N_k \ln\theta_k$ with the regularisers of the OT problem. In general, as $N_k = n\theta_k \left(1 + \mathrm{O}_p\big(\sqrt{(1-\theta_k)/(n\theta_k)}\big)\right)$

$$\sum_{k=1}^{K} N_k \log\theta_k = n \sum_{k=1}^{K} \theta_k \log\theta_k + \mathrm{O}_p(\sqrt{n}).$$

Remark that

$$\begin{aligned}\|\pi\|_F^2 &= \sum_{1\leq i,j\leq n}\left(\sum_{k=1}^{K} Z_{ik} Z_{jk} \mathbb{1}_{i\neq j} (N_k - 1)^{-1}\right)^2 = \sum_{1\leq i,j\leq n}\sum_{k=1}^{K} Z_{ik} Z_{jk} \mathbb{1}_{i\neq j} (N_k - 1)^{-2} \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik} (N_k - 1)^{-1} = \sum_{k=1}^{K} N_k (N_k - 1)^{-1} = K + \mathrm{O}_p\left(\sum_{k=1}^{K} n^{-1}\theta_k^{-1}\right).\end{aligned}$$

Therefore, while maximising the approximate surrogate likelihood is the maximisation of

$$\pi \mapsto n \sum_k \theta_k \log \theta_k + \mathrm{o}_p(n) - \frac{1}{\sigma^2}\langle \pi, C\rangle - nd(\log\sigma - \sigma^2),$$

the quadratically regularised optimal transport involves minimisation of

$$\pi \mapsto \langle \pi, C\rangle_F + \varepsilon \|\pi\|_F^2$$

over the set of hollow bistochastic matrices. In view of the developments above, one can can rewrite the first minimisation problem, up to asymptotically negligible factors, as

$$\max_{\pi \in B_n^\circ} \frac{-nH(\theta)\|\pi\|_F^2}{K} - \frac{1}{\sigma^2}\langle \pi, C\rangle_F - nd(\log\sigma - \sigma^2),$$

Hence, the above computations suggest the choice

$$\varepsilon = \frac{n\sigma^2}{K} H(\theta)$$

to make both problems nearly equivalent as σ^2, θ and K are assumedly known and fixed. The claim of Theorem 1 thus follows.

3.2 Robustness

It remains to provide one final proposition.

Proposition 3. *Assume that Y_i are as above, i.e., independent and normally distributed with covariance matrix $\sigma^2 I_d$. Only the mean vector depends on the cluster. Assume that there exist a constant C_μ such that $\max_{1\le k,k'\le K} \|\mu_k - \mu_{k'}\| \le C_\mu$. Now consider corrupted observations $\tilde{Y}_i = Y_i + \eta_i$, with $\eta_i \sim \mathcal{N}(0, \Sigma_i)$. The $\Sigma_i's$ can be different as long as $\|\Sigma_i\|_2 \le C_\eta, \forall i$. Then,*

$$\|\tilde{Y}_i - \tilde{Y}_j\|^2 = \|Y_i - Y_j\|^2 + \|\eta_i\|^2 + \|\eta_j\|^2 + \mathrm{O}_p\left(\sqrt{d}\right).$$

Proof. The proof is a mere computation of moments and classical trace bounds, which we leave to the reader for space considerations. □

In words, the previous proposition states that the cost matrix of the perturbed points—giving rise to anisotropic Gaussian mixtures, is equal to that of the isotropic case up to a controlled error and a rank-one matrix.

4 Further Remarks and Conclusion

A regularisation with the Kullback–Leibler even though possible is less adapted to the problem. Indeed, contrarily to the case of the quadratic regularisation, as the optimal coupling is not sparse, one cannot expect an entropy regularised transport plan to be close to the matrix π defined above and depending on the Z's. It is also natural to wonder whether the approach works for non-Gaussian mixture models. The answer is not obvious and given the good results of general pseudo-likelihood estimators in certain cases, one might hope for some stability in the case of slight model misspecification. Another possibility would be to consider transformations

of the distance matrix that mimic the log-likelihood of the mixture components. It seems unlikely that other models will give rise to the same nice properties as in the Gaussian case.

In conclusion, we have shown a link between quadratic-regularised (hollow) self-transport and maximum likelihood estimation in Gaussian mixture models. Also, the choice of hollow bistochastic matrices is important for both approximation and robustness reasons. This links shows that applying spectral clustering techniques on the obtained transport plan is a statistically principled approach. The method is also easy to apply as it only requires choosing one parameter. This also suggests that QOT might be used to complement the current EM algorithms or help find a good initialisation for other methods. As the spectrum of π^* contains relevant information, it might be used to refine other statistical procedures.

Acknowledgements. Gilles Mordant acknowledges funding of the DFG within CRC1456- A04.

References

1. Doss, N., Wu, Y., Yang, P., Zhou, H.H.: Optimal estimation of high-dimensional Gaussian location mixtures. Ann. Stat. **51**(1), 62–95 (2023)
2. Landa, B., Cheng, X.: Robust inference of manifold density and geometry by doubly stochastic scaling. SIAM J. Math. Data Sci. **5**(3), 589–614 (2023)
3. Landa, B., Coifman, R.R., Kluger, Y.: Doubly stochastic normalization of the Gaussian kernel is robust to heteroskedastic noise. SIAM J. Math. Data Sci. **3**(1), 388–413 (2021)
4. Lim, D., Vidal, R., Haeffele, B.D.: Doubly stochastic subspace clustering. ArXiv preprint arXiv:2011.14859 (2020)
5. Löffler, M., Zhang, A.Y., Zhou, H.H.: Optimality of spectral clustering in the Gaussian mixture model. Ann. Stat. **49**(5), 2506–2530 (2021)
6. Marshall, N.F., Coifman, R.R.: Manifold learning with bi-stochastic kernels. IMA J. Appl. Math. **84**(3), 455–482 (2019)
7. Villani, C.: Optimal Transport: Old and New. A Series of Modern Surveys in Mathematics, Vol. 338. Springer, Berlin (2009)
8. Wu, Y., Zhou, H.H.: Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. Math. Stat. Learn. **4**(3), 143–220 (2021)
9. Zhang, S., Mordant, G., Matsumoto, T., Schiebinger, G.: Manifold learning with sparse regularised optimal transport. ArXiv preprint arXiv:2307.09816 (2023)

Stochastic Orders Under Uncertainty

Alfred Müller(✉)

Universität Siegen, Department Mathematik, 57072 Siegen, Germany
mueller@mathematik.uni-siegen.de

Abstract. We study stochastic order relations under uncertainty. This topic has hardly been considered in the literature so far. It is shown that we cannot expect any reasonable robustness of classical first order stochastic dominance, whereas such properties hold for some versions of almost stochastic dominance, as almost stochastic dominance is naturally related to robustness with respect to the Wasserstein distance. We also review other recent results on almost stochastic dominance under uncertainty where the uncertainty sets are either described by knowing only mean and variance or by knowing only the marginal distribution in a multivariate context.

Keywords: Stochastic order · Uncertainty · Robustness · Almost stochastic dominance

1 Introduction

Decision making under risk is a very difficult task. It requires knowledge of the distributions of the risks under consideration *and* knowledge of the preference relations of the decision maker. Both are often subject to uncertainty or difficult to assess, in particular in multivariate settings. Sufficient conditions for unambiguous decisions under only partial knowledge of the preferences is the topic of *stochastic dominance* or *stochastic orders*. This is a widely studied topic in probability theory as well as in applications. We refer to the monographs of Müller and Stoyan [3] and Shaked and Shanthikumar [8] for comprehensive studies on that subject. There is also a large literature on the effect of uncertainty of distributions on quantities of interest, e.g. assuming only knowledge of some moments or only the marginals in a multivariate context. For a recent monograph on how to handle such model risk we refer to Rüschendorf et al. [7]. In this paper we want to combine the two topics studying stochastic orders under uncertainty.

2 Stochastic Orders: Preliminaries

Throughout the paper we will assume that all random quantities are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and that all mentioned expectations exist. Given real

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 302–308, 2024.
https://doi.org/10.1007/978-3-031-65993-5_37

random variables X and Y we will denote the cumulative distribution functions by

$$F(t) := \mathbb{P}(X \le t),\ G(t) := \mathbb{P}(Y \le t),\ t \in \mathbb{R}.$$

We will also frequently use the generalized inverses defined as

$$F^{-1}(u) := \inf\{t \in \mathbb{R} : F(t) \ge u\},\ u \in (0,1).$$

Given a class $\mathcal{U}$ of functions $u : \mathbb{R} \to \mathbb{R}$ and two random variables X and Y, we say that $X \le_{\mathcal{U}} Y$ if $\mathbb{E}u(X) \le \mathbb{E}u(Y)$ for all $u \in \mathcal{U}$, whenever both expectations exist. If $\mathcal{U}$ consists of all increasing functions, the we say that X is dominated by Y in the usual stochastic order. Using the notation of Müller and Stoyan [3] and Shaked and Shanthikumar [8], this relation is denoted by $X \le_{st} Y$. In this case, we also write $F \le_{st} G$. The condition is equivalent to pointwise ordering of the corresponding distribution functions or inverse distribution functions:

$$X \le_{st} Y \iff F(t) \ge G(t) \text{ for all } t \in \mathbb{R} \iff F^{-1}(u) \le G^{-1}(u) \text{ for all } u \in (0,1). \quad (1)$$

Leshno and Levy [2] introduced a version of almost stochastic dominance for distributions with bounded support. This has been generalized in Müller et al. [4,5] and we adopt the following definition and results from there.

Definition 1. *For a given number $\gamma \in [0,1]$, the symbol $\mathcal{U}_\gamma$ denotes the set of functions u such that we have*

$$0 \le \gamma u'(x) \le u'(z) \text{ for all } x, z \in \mathbb{R}. \quad (2)$$

We say that $X \le_\gamma Y$ if $\mathbb{E}u(X) \le \mathbb{E}u(Y)$ for all $u \in \mathcal{U}_\gamma$.

For arbitrary random variables X and Y with finite means and distributions F and G we get $X \le_\gamma Y$ if and only if

$$\int_{-\infty}^{\infty} (G(y) - F(y))_+ \, \mathrm{d}y \le \gamma \int_{-\infty}^{\infty} (F(y) - G(y))_+ \, \mathrm{d}y, \quad (3)$$

where $(y)_+ := \max(y, 0)$. Equivalently,

$$\int_0^1 (F^{-1}(y) - G^{-1}(y))_+ \, \mathrm{d}y \le \gamma \int_0^1 (G^{-1}(y) - F^{-1}(y))_+ \, \mathrm{d}y, \quad (4)$$

Notice that this parametric family of stochastic orders gets weaker and weaker for increasing γ and interpolates between the case of $X \le_{st} Y$ for $\gamma = 0$ and just comparing the means $\mathbb{E}X \le \mathbb{E}Y$ for $\gamma = 1$.

3 Stochastic Orders: Uncertainty and Robustness

We want to consider now the question of stochastic ordering, if we have only partial knowledge of the distribution. Assume we have an arbitrary stochastic order $\preceq$. Given two classes of distributions $\mathcal{F}$ and $\mathcal{G}$ we would be interested in situations where we have $F \preceq G$ for all $F \in \mathcal{F}$ and all $G \in \mathcal{G}$. Of particular interest are classes of distributions of the following types.

– We may only have knowledge of the values μ_i of a finite number of functionals ρ_i of the random variables, i.e.

$$\mathcal{F} := \{F_X : \rho_i(X) = \mu_i,\ i = 1, \dots, n\}.$$

The most famous example is that we may only know a few first moments, e.g. only mean and variance. We will consider this case later.
– In the case of multivariate distributions an interesting type of uncertainty is given by *dependence uncertainty*, where we only know the marginals $F_1, \dots, F_d$ of a random vector. This class of distributions is typically called a Fréchet class. This case will also be considered later.
– When talking about robustness one typically assumes that one has an estimated distribution F_0 and considers all distributions that differ at most by some ε with respect to some distance d for distribution functions, i.e.

$$\mathcal{F} := \{F : d(F, F_0) \le \varepsilon\}.$$

As the most important example of such a distance we will consider here the Wasserstein distance (of order 1)

$$W_1(F, G) := \int_{-\infty}^{\infty} |G(y) - F(y)| \, \mathrm{d}y = \int_0^1 |G^{-1}(y) - F^{-1}(y)| \, \mathrm{d}y$$

which is known to be also the solution of the optimal transportation problem

$$W_1(F, G) := \inf\{\mathbb{E}|X - Y| : X \sim F, Y \sim G\}.$$

Unfortunately, it can be derived from the results in Durante et al. [1] that $\le_{st}$ as well as weaker stochastic dominance relations as second order stochastic dominance do not have any robustness in that sense. They show Baire category results, which can be restated as follows: for any bounded distributions F and G and any $\varepsilon > 0$ we can find F_ε and G_ε with $W_1(F_\varepsilon, F) \le \varepsilon$ and $W_1(G_\varepsilon, G) \le \varepsilon$ such that $F_\varepsilon \not\le_{st} G_\varepsilon$. This result can be generalized as follows.

Theorem 1. *For any distributions F and G with finite means and any $\varepsilon > 0$ we can find a distribution F_ε with $W_1(F_\varepsilon, F) \le \varepsilon$ such that $F_\varepsilon \not\le_{st} G$.*

Proof. Fix $\varepsilon > 0$ and choose some $a < b$ such that $\delta := F(b) - F(a) > 0$. Define $\delta' := \min\{\delta, \varepsilon/(b-a)\}$. For $t > \delta'$ we define the cumulative distribution function

$$F_t(x) := \begin{cases} F(x), & x < a, \\ \min\{F(x), F(b) - \delta'/t\}, & a \le x < b, \\ F(x) - \delta'/t, & b \le x < c, \\ F(x), & x \ge c, \end{cases}$$

where c is chosen such that $W_1(F, F_t) = \varepsilon$. Notice that the Wasserstein distance between these two distributions is linearly increasing in the parameter c with

$W_1(F, F_t) < \varepsilon$ for $c = b$ so that we always find a unique $c > b$ such that $W_1(F, F_t) = \varepsilon$. Morevoer, we then have

$$\varepsilon < \frac{\delta'}{t} \cdot (c - a), \quad \text{thus } c > a + \frac{\varepsilon t}{\delta'}.$$

Therefore $F_t(a + \varepsilon t/\delta') \leq 1 - t/\delta'$. Now assume that $F_t \leq_{st} G$ for all $t > \delta'$. Then we must have $G(a + \varepsilon t/\delta') \leq 1 - t/\delta'$ for all $t > \delta'$ and thus we get

$$\int_0^\infty (1 - G(x))\,\mathrm{d}x = \infty,$$

which is a contradiction to the finiteness of the mean of G.

This result clearly shows that we cannot expect any robustness properties for the usual stochastic ordering $\leq_{st}$. If we want robustness results we have to look for weaker orderings and almost stochastic dominance as introduced in Definition 1 is the most natural candidate for robustness results with respect to the Wasserstein distance. There is a natural close relationship between the concept of almost stochastic dominance and the Wasserstein distance W_1 as both can be characterized by the areas between distribution functions or their inverses. Indeed, let us define for arbitrary distributions F and G with finite means μ_F and μ_G the quantity

$$D^+(F, G) := \int_{-\infty}^{\infty} (G(y) - F(y))_+\,\mathrm{d}y \;=\; \int_0^1 (F^{-1}(y) - G^{-1}(y))_+\,\mathrm{d}y. \tag{5}$$

Then the following relationships hold.

$$F \leq_\gamma G \quad \text{if and only if} \quad \gamma \geq \gamma^*(F, G) := \frac{D^+(F, G)}{D^+(G, F)}, \tag{6}$$

$$W_1(F, G) = D^+(F, G) + D^+(G, F), \tag{7}$$

and for the difference of the means we get

$$\mu_G - \mu_F = \int_0^1 (G^{-1}(y) - F^{-1}(y))\,\mathrm{d}y \;=\; D^+(G, F) - D^+(F, G), \tag{8}$$

Using these identities we can derive the following robustness result for almost stochastic dominance.

Theorem 2. *Assume that F and G are distributions with finite means and $\delta := \mu_G - \mu_F > 0$. Then $\gamma^* := \gamma^*(F, G) < 1$ and $F \leq_\gamma G$ for all $\gamma \geq \gamma^*$. If $W_1(F, F') \leq \varepsilon$, then*

$$\gamma^*(F', G) \leq \frac{\gamma^*\delta + (1 - \gamma^*)\varepsilon}{\delta - (1 - \gamma^*)\varepsilon}. \tag{9}$$

In particular, if $F \leq_{st} G$, then $F' \leq_\gamma G$ for all $\gamma \geq \varepsilon/(\delta - \varepsilon)$.

Proof. The following triangle inequalities holds for D^+:

$$D^+(F',G) \le D^+(F',F) + D^+(F,G) \text{ and } D^+(G,F) \le D^+(G,F') + D^+(F',F). \tag{10}$$

As $W_1(F,F') = D^+(F,F') + D^+(F',F) \le \varepsilon$ this implies

$$\frac{D^+(F',G)}{D^+(G,F')} \le \frac{D^+(F',F) + D^+(F,G)}{D^+(G,F) - D^+(F',F)} \le \frac{D^+(F,G) + \varepsilon}{D^+(G,F) - \varepsilon}. \tag{11}$$

From (6) and (8) we get for $\delta := \mu_G - \mu_F > 0$

$$D^+(F,G) = \frac{\gamma^*\delta}{1-\gamma^*} \text{ and } D^+(G,F) = \frac{\delta}{1-\gamma^*}$$

Inserting this into (11) yields

$$\frac{D^+(F',G)}{D^+(G,F')} \le \frac{\gamma^*\delta + (1-\gamma^*)\varepsilon}{\delta - (1-\gamma^*)\varepsilon} \tag{12}$$

and thus (9). The claim for $\le_{st}$ follows for the case $\gamma^* = 0$.

Almost stochastic dominance results when knowing only means and variances have been considered in Müller et al. [5]. Müller et al. [6] have considered multivariate versions of almost stochastic dominance under dependence uncertainty. We restate some of the basic results given there in the notation used here.

Theorem 3 (Müller et al. [5]). *Let X and Y be random variables with $\mathbb{E}(X) = \mu_X$, $Var(X) = \sigma_X^2$, $\mathbb{E}(Y) = \mu_Y$, $Var(X) = \sigma_Y^2$, and $\mu_Y > \mu_X$. Define*

$$t = \frac{\mu_Y - \mu_X}{\sigma_Y + \sigma_X}. \tag{13}$$

Then $X \le_\gamma Y$ for

$$\gamma \ge \gamma^* = \frac{1}{1 + 2t\left(t + \sqrt{t^2+1}\right)}. \tag{14}$$

To state the results for dependence uncertainty we first have to define multivariate versions of almost stochastic dominance, which are natural extensions of the univariate case.

Given two random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$, the following notation for their joint distribution functions will be used:

$$F(t_1,\ldots,t_d) := \mathbb{P}(X_1 \le t_1, \ldots, X_d \le t_d),\ G(t_1,\ldots,t_d) := \mathbb{P}(Y_1 \le t_1, \ldots, Y_d \le t_d),$$

with $t_1, \ldots, t_d \in \mathbb{R}$. The corresponding notation for their marginal distribution functions is:

$$F_i(t_i) := \mathbb{P}(X_i \le t_i),\ G_i(t_i) := \mathbb{P}(Y_i \le t_i),\ i = 1, \ldots, d.$$

For partial derivatives of functions $u : \mathbb{R}^d \to \mathbb{R}$ we write

$$u_i'(\boldsymbol{x}) := \frac{\partial u(\boldsymbol{x})}{\partial x_i}.$$

Definition 2 (Müller et al. [6]). *For a given vector* $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_d) \in [0,1]^d$, *the symbol* $\mathcal{U}_{\boldsymbol{\gamma}}$ *denotes the set of functions such that, for all* $i \in \{1, \ldots, d\}$, *we have*

$$0 \leq \gamma_i u_i'(\boldsymbol{y}) \leq u_i'(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d. \tag{15}$$

Notice that the condition in inequality (15) *is equivalent to*

$$\frac{\inf u_i'(\boldsymbol{x})}{\sup u_i'(\boldsymbol{x})} \geq \gamma_i. \tag{16}$$

We say that $\boldsymbol{X} \leq_{\boldsymbol{\gamma}} \boldsymbol{Y}$ *holds if*

$$\mathbb{E}u(\boldsymbol{X}) \leq \mathbb{E}u(\boldsymbol{Y}) \text{ for all } u \in \mathcal{U}_{\boldsymbol{\gamma}}.$$

Then we can state the following result about dependence uncertainty.

Theorem 4 (Müller et al. [6]). *Assume that the marginal distributions of the components of* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are known and that* $\mu_{X_i} \leq \mu_{Y_i}$ *for all* $i = 1, \ldots, d$. *Let* $\delta_i := \inf\{x : F_i(x) + G_i(x) \geq 1\}$ *and let*

$$\gamma_i := \frac{\mathbb{E}[(\delta_i - Y_i)_+] + \mathbb{E}[(X_i - \delta_i)_+]}{\mathbb{E}[(Y_i - \delta_i)_+] + \mathbb{E}[(\delta_i - X_i)_+]}, \tag{17}$$

for $i = 1, \ldots, d$. *Then* $\boldsymbol{X} \leq_{\boldsymbol{\gamma}} \boldsymbol{Y}$.

For the case that we only know the means and variances of the marginals we get the following result.

Define

$$\zeta(t) := \frac{1}{1 + 2t(t + \sqrt{t^2 + 1})}. \tag{18}$$

Theorem 5 (Müller et al. [6]). *Let the two random vectors* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *have finite means and variances. Moreover, for all* $i = 1, \ldots, d$, *let* $\mu_{X_i} \leq \mu_{Y_i}$ *and let*

$$\tau_i = \frac{\mu_{Y_i} - \mu_{X_i}}{\sigma_{X_i} + \sigma_{Y_i}}.$$

If $\gamma_i = \zeta(\tau_i)$, $i = 1, \ldots, d$, *then* $\boldsymbol{X} \leq_{\boldsymbol{\gamma}} \boldsymbol{Y}$.

For a second variant of multivariate almost stochastic dominance we consider a parameter $\gamma \in [0,1]$ and parameters $\beta_1, \ldots, \beta_d$ describing the relevance of the components such that as in the univariate case we get a complete order for $\gamma = 1$ which holds if and only if

$$\mathbb{E}\left[\sum_{i=1}^{d} \beta_i X_i\right] \leq \mathbb{E}\left[\sum_{i=1}^{d} \beta_i Y_i\right].$$

Definition 3 (Müller et al. [6]). *For $\gamma \in [0,1]$ and $\boldsymbol{\beta} \in \mathbb{R}^d_+$, let $\mathcal{U}_{\gamma,\boldsymbol{\beta}}$ be the class of functions $u\colon \mathbb{R}^d \to \mathbb{R}$ such that*

$$0 < \gamma\beta_i \le u_i'(\boldsymbol{x}) \le \beta_i \quad \text{for all } i \in \{1,\dots,d\}. \tag{19}$$

The random vector $\boldsymbol{X}$ is dominated by the random vector $\boldsymbol{Y}$ in the sense of $(\gamma,\boldsymbol{\beta})$-MASD ($\boldsymbol{X} \le_{\gamma,\boldsymbol{\beta}} \boldsymbol{Y}$) if

$$E[u(\boldsymbol{X})] \le E[u(\boldsymbol{Y})] \quad \text{for all } u \in \mathcal{U}_{\gamma,\boldsymbol{\beta}}.$$

For that case we can derive similar results:

Theorem 6 (Müller et al. [6]). *Assume that the marginal distributions of the components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are known. Let $\delta_i := \inf\{x\colon F_i(x)+G_i(x) \ge 1\}$ and let*

$$\gamma := \frac{\sum_{i=1}^d \beta_i(\mathbb{E}[(\delta_i - Y_i)_+] + \mathbb{E}[(X_i - \delta_i)_+])}{\sum_{i=1}^d \beta_i(\mathbb{E}[(Y_i - \delta_i)_+] + \mathbb{E}[(\delta_i - X_i)_+])}.$$

If

$$\sum_{i=1}^d \beta_i \mu_{X_i} \le \sum_{i=1}^d \beta_i \mu_{Y_i}, \tag{20}$$

then $\boldsymbol{X} \le_{\gamma,\boldsymbol{\beta}} \boldsymbol{Y}$.

These results show that it is possible to show versions of stochastic dominance under uncertainty, if one uses the right concepts of almost stochastic dominance. We hope that this stipulates further research on the topic.

References

1. Durante, F., Fernández-Sánchez, J., Ignazzi, C.: Baire category results for stochastic orders. Rev. Real Acad. Cienc. Exactas Fis. Nat. A: Mat. **116**(4), 188 (2022)
2. Leshno, M., Levy, H.: Preferred by "all" and preferred by "most" decision makers: almost stochastic dominance. Management Sci. **48**(8), 1074–1085 (2002). http://dx.doi.org/10.1287/mnsc.48.8.1074.169
3. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. Wiley, Chichester (2002)
4. Müller, A., Scarsini, M., Tsetlin, I., Winkler, R.L.: Between first and second-order stochastic dominance. Management Sci. **63**(9), 2933–2947 (2017). http://dx.doi.org/10.1287/mnsc.2016.2486
5. Müller, A., Scarsini, M., Tsetlin, I., Winkler, R.L.: Ranking distributions when only means and variances are known. Oper. Res. **70**, 2851–2859 (2022). https://doi.org/10.1287/opre.2020.2072
6. Müller, A., Scarsini, M., Tsetlin, I., Winkler, R.L.: Multivariate almost stochastic dominance: Transfer characterizations and sufficient conditions under dependence uncertainty. Oper. Res. (2023). https://pubsonline.informs.org/doi/10.1287/opre.2022.0596
7. Rüschendorf, L., Vanduffel, S., Bernard, C.: Model Risk Management: Risk Bounds under Uncertainty. Cambridge University Press (2023)
8. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer, New York (2007). https://doi.org/10.1007/978-0-387-34675-5

Conditions for the Multivariate Stochastic Order Under Dependence Uncertainty

Alfred Müller[1] and Marco Scarsini[2(✉)]

[1] Department Mathematik, Universität Siegen, 57072 Siegen, Germany
mueller@mathematik.uni-siegen.de
[2] Dipartimento di Economia e Finanza, Luiss University, 00197 Roma, Italy
marco.scarsini@luiss.it

Abstract. In the univariate case checking the usual stochastic order for two random variables is equivalent to comparing point-wise their distribution functions.Therefore, the task is easy.The multivariate case this is much more complicated to handle, because it requires the comparison of the probability of a huge class of sets (all the upper sets). In this paper we consider conditions for multivariate stochastic order for the case that we only know the marginal distributions and have no or only partial knowledge of the dependence structure.

Keywords: multivariate stochastic order · sufficient conditions · dependence uncertainty

1 Introduction

Comparing two univariate random variables in the sense of the usual stochastic order amounts to a pointwise comparison of their distribution functions. Comparing two random vectors requires to compare the probabilities that they lie in the same upper set, for all possible upper sets. Even upon discretization of the random vectors, this is an extremely hard endeavor.

It is well known that, even if all the components of a random vector $\boldsymbol{X}$ are stochastically dominated by the corresponding components of $\boldsymbol{Y}$, the vector $\boldsymbol{X}$ is not necessarily stochastically dominated by $\boldsymbol{Y}$. The dependence structures of $\boldsymbol{X}$ and $\boldsymbol{Y}$ play a major role in the domination or lack thereof. On the other hand, it is often difficult to elicit joint distribution functions, even when their marginals are completely known. These situations go under the name of *dependence uncertainty*.

The main goal of this paper is to deal with necessary and sufficient conditions for multivariate stochastic order of $\boldsymbol{Y}$ with respect to $\boldsymbol{X}$ for the case that we only know the marginal distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ and have no knowledge of the dependence structure.

Among other things, we prove that, under complete dependence uncertainty, the support of the dominated vector must have a finite upper bound and the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 309–316, 2024.
https://doi.org/10.1007/978-3-031-65993-5_38

support of the dominant vector must have a finite lower bound. When the two vectors are bounded, some non-trivial sufficient conditions can be found that impose a limited overlap for the support of the two random vectors.

Some papers have dealt with the interplay between marginal stochastic orders, dependence structure of random vectors, and multivariate stochastic orders. For instance, Rüschendorf [9] and Scarsini [10] proved that if two random vectors have the same copula, then the stochastic order of the marginals implies the multivariate stochastic order of the vectors. Scarsini [11] showed that a similar implication does not hold for the convex order of random vectors. Müller and Scarsini [4] proved that if two random vectors have the same stochastically increasing copula, then the convex order of the marginals implies the directionally convex order of the vectors. Müller [3] proved that two random vectors having multivariate normal distributions can be stochastically ordered only if they have the same covariance matrix. Yin [14] extended this result to elliptically contoured distributions. Müller et al. [6] determined conditions for marginal dominance that implies a version of multivariate almost stochastic dominance under dependence uncertainty.

2 Interplay Between Dependence and Marginal Dominance

Given two random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$, the following notation for their joint distribution function (JDFs) will be used throughout the paper:

$$F(t_1, \ldots, t_d) := \mathbb{P}(X_1 \le t_1, \ldots, X_d \le t_d),\ G(t_1, \ldots, t_d) := \mathbb{P}(Y_1 \le t_1, \ldots, Y_d \le t_d),$$

with $t_1, \ldots, t_d \in \mathbb{R}$. The corresponding notation for their marginal distribution functions is:

$$F_i(t_i) := \mathbb{P}(X_i \le t_i),\ G_i(t_i) := \mathbb{P}(Y_i \le t_i),\ i = 1, \ldots, d.$$

The symbol 1_A indicates the indicator of set A.

Given a class $\mathcal{U}$ of functions $u : \mathbb{R}^d \to \mathbb{R}$ and two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we say that $\boldsymbol{X} \le_{\mathcal{U}} \boldsymbol{Y}$ if $\mathbb{E}[u(\boldsymbol{X})] \le \mathbb{E}[u(\boldsymbol{Y})]$ for all $u \in \mathcal{U}$, whenever both expectations exist. If $\mathcal{U}$ consists of all increasing functions, the we say that $\boldsymbol{X}$ is dominated by $\boldsymbol{Y}$ in the *multivariate stochastic order* (MSO). The concept of multivariate stochastic order was first studied in the statistics literature by Lehmann [2]. Using the notation of Müller and Stoyan [5] and Shaked and Shanthikumar [12], this relation is denoted by $\boldsymbol{X} \le_{st} \boldsymbol{Y}$. In this case, we also write $F \le_{st} G$.

In the univariate case the condition for $\le_{st}$ is equivalent to pointwise ordering of the corresponding distribution functions:

$$X \le_{st} Y \iff F(t) \ge G(t) \quad \text{for all } t \in \mathbb{R}. \tag{1}$$

If the distributions have densities with respect to some dominating measure, some easy-to-check conditions for dominance exist. For instance if the densities

cross exactly once, then $\leq_{st}$ holds between the two distributions. In contrast, if they cross exactly twice, then $\leq_{st}$ cannot hold.

Unfortunately, in the multivariate case the situation is much more complex: the condition $F_{\boldsymbol{X}}(\boldsymbol{t}) \geq F_{\boldsymbol{Y}}(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathbb{R}^d$ is only necessary for $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$. Moreover, even this condition is typically not easy to verify. The characterization of MSO corresponding to (1) becomes

$$\boldsymbol{X} \leq_{st} \boldsymbol{Y} \iff \mathbb{P}(\boldsymbol{X} \in A) \leq \mathbb{P}(\boldsymbol{X} \in A) \quad \text{for every upper set } A \subset \mathbb{R}^d, \tag{2}$$

where a set $A \subset \mathbb{R}^d$ is called *upper* if $\boldsymbol{x} \in A$ and $\boldsymbol{x} \leq \boldsymbol{y}$ imply $\boldsymbol{y} \in A$ (see, e.g. [1]). This condition is hard to verify. For discrete distributions it is possible to use a characterization via transfers, from which a transportation problem can be derived and solved (see [7] for details). Even this case is not easy to deal with.

Notice that $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$ implies $X_i \leq_{st} Y_i$ for all $i = 1, \ldots, d$, but the converse implication is false. This is due to the fact that the dependence within the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ has a relevant role for their dominance conditions. Sklar [13] proved that, for every d-variate distribution function F with marginals $F_1, \ldots, F_d$, there exists a function $C : [0,1]^d \to [0,1]$, called *copula*, such that

$$F(t_1, \ldots, t_d) = C(F_1(t_1), \ldots, F_d(t_d)), \quad \text{for all } (t_1, \ldots, t_d) \in \mathbb{R}^d. \tag{3}$$

The copula is a JDF with uniform marginals on $[0,1]$; the function C describes the dependence structure of the random vector. The following theorem shows that, when two random vectors have the same copula, then they are stochastically ordered if and only if their marginals are.

Theorem 1 (Rüschendorf [9], Scarsini [10]). *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ have a common copula. Then $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$ if and only if $X_i \leq_{st} Y_i$ for all $i = 1, 2, \ldots, d$.*

Equality of the two copulas is a quite restrictive condition. For some parametric families of distributions, even more restrictive conditions are required, as the following theorem shows. In the following theorem the symbol $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Theorem 2 (Müller [3]). *Let $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}})$ and $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}})$. Then $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$ if and only if $\mu_i \leq \mu_i'$ for all $i = 1, ..., d$, and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Sigma}_{\boldsymbol{Y}}$.*

In practical applications the parameters in the covariance matrix $\boldsymbol{\Sigma}$ are typically estimated, so it is almost impossible to get exactly the same values. Therefore dominance conditions for the important case of multivariate normal distributions are almost never satisfied. Notice that this is true even when the difference between the means of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is large. So, a small difference in the copula can have the effect that $\boldsymbol{X} \not\leq_{st} \boldsymbol{Y}$ even if marginals of $\boldsymbol{Y}$ are obtained by a large shift from the marginals of $\boldsymbol{X}$. We now show that this is not a peculiar feature of the multivariate normal distribution. The following example demonstrates that this effect appears for any distribution with exponentially decaying tails.

Example 1. Consider a two-dimensional vector $\boldsymbol{X} = (X_1, X_2)$ whose components have the same marginal distribution $F(x) = \mathbb{P}(X_1 \le x) = \mathbb{P}(X_2 \le x)$. Let the marginal distributions of $\boldsymbol{Y} = (Y_1, Y_2)$ be such that $\mathbb{P}(Y_1 \le x) = \mathbb{P}(Y_2 \le x) = F(x-t)$, i.e., Y_i has the same distribution as $X_i + t$ for some $t > 0$. Moreover, assume that the distributions have exponentially decaying tails, i.e., we assume that for some $\alpha > 0$ the moment generating function $\hat{m}_X(z) := \mathbb{E}(e^{zX})$ is finite for $0 \le z < \alpha$ and infinite for $z > \alpha$. This implies that for independent random variables Y_1 and Y_2 with this property, the same property holds for the sum $Y_1 + Y_2$, whereas, for the sum of two identical random variables $X_1 = X_2$, we get $\mathbb{E}(e^{z(X_1+X_2)}) = \infty$ for $z > \alpha/2$. Now assume that Y_1 and Y_2 are independent, whereas the joint distribution of (X_1, X_2) is a mixture of independence and comonotonicity, i.e.,

$$\mathbb{P}(X_1 \le t_1, X_2 \le t_2) = (1-\varepsilon)F(t_1)F(t_2) + \varepsilon \min\{F(t_1), F(t_2)\} \tag{4}$$

for some $\varepsilon \in (0,1)$. This implies for the increasing function $f(x_1, x_2) = e^{z(x_1+x_2)}$ with $\alpha/2 < z < \alpha$ that $\mathbb{E}(f(Y_1, Y_2)) < \infty = \mathbb{E}(f(X_1, X_2))$ and hence $\boldsymbol{X} \not\le_{st} \boldsymbol{Y}$ even if t is very large and ε very small. Notice that a small ε means that the copulas are almost the same as we have $\|C_{\boldsymbol{X}} - C_{\boldsymbol{Y}}\|_\infty \le \varepsilon$.

In Example 1 $\boldsymbol{X} \not\le_{st} \boldsymbol{Y}$ because of the tail behavior of the random variables, which translates into the behavior of the copula at the boundary of its domain. In contrast, the behavior of the copula at the center of its domain does not affect the relation $\boldsymbol{X} \le_{st} \boldsymbol{Y}$ if the marginals are substantially different. We now show this when there is a shift of the marginals.

Example 2. Consider a random vector $\boldsymbol{X}$ with marginals $F_i,\ i = 1, \dots, d$, and copula $C_{\boldsymbol{X}}$. Let m_i denote the median of $X_i,\ i = 1, \dots, d$. Assume that the random vector $\boldsymbol{Y}$ has a copula $C_{\boldsymbol{Y}}$ and marginals $G_i,\ i = 1, \dots, d$, where G_i is obtained from F_i via a shift by an amount $a_i > 0$, i.e., $G_i(t) = F_i(t - a_i),\ t \in \mathbb{R}$. This implies that $G_i^{-1}(u) = F_i^{-1}(u) + a_i,\ u \in (0,1)$. We define the set

$$A := \left\{ \boldsymbol{x} \in [0,1]^d : F_i\left(m_i - \frac{a_i}{2}\right) \le x_i \le F_i\left(m_i + \frac{a_i}{2}\right),\ i = 1, \dots, d \right\}. \tag{5}$$

We claim that if $C_{\boldsymbol{X}}(\boldsymbol{u}) = C_{\boldsymbol{Y}}(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d \setminus A$, then $\boldsymbol{X} \le_{st} \boldsymbol{Y}$, independently of the values of $C_{\boldsymbol{X}}(\boldsymbol{u})$ and $C_{\boldsymbol{Y}}(\boldsymbol{u})$ for $\boldsymbol{u} \in A$. This can be shown as follows.

Let $\boldsymbol{U}$ be a random vector with values in $(0,1)^d$ with uniform marginals and distribution function $C_{\boldsymbol{X}}$. If $C_{\boldsymbol{X}}(\boldsymbol{u}) = C_{\boldsymbol{Y}}(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d \setminus A$, then we can find a random vector $\boldsymbol{V}$ with distribution function $C_{\boldsymbol{Y}}$ such that $\boldsymbol{V} = \boldsymbol{U}$ if $\boldsymbol{U} \in (0,1)^d \setminus A$. Consider $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ derived from $\boldsymbol{U}$ and $\boldsymbol{V}$ via the standard construction

$$\boldsymbol{X}' = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)), \quad \boldsymbol{Y}' = (G_1^{-1}(V_1), \dots, G_d^{-1}(V_d)). \tag{6}$$

When $\boldsymbol{U} \in [0,1]^d \setminus A$, we have $\boldsymbol{V} = \boldsymbol{U}$ and $\boldsymbol{Y}' = \boldsymbol{X}' + \boldsymbol{a} \ge \boldsymbol{X}'$. On the other hand, if $\boldsymbol{U} \in A$, then

$$F_i\left(m_i - \frac{a_i}{2}\right) \le U_i \le F_i\left(m_i + \frac{a_i}{2}\right). \text{ Hence, } m_i - \frac{a_i}{2} \le F_i^{-1}(U_i) \le m_i + \frac{a_i}{2},\ i = 1, \dots, d. \tag{7}$$

As $G_i^{-1}(u_i) = F_i^{-1}(u_i) + a_i$ we get $G_i^{-1}(V_i) \geq m_i + a_i/2 \geq F_i^{-1}(U_i)$. Hence, $\boldsymbol{Y}' \geq \boldsymbol{X}'$ if $\boldsymbol{U} \in A$. Therefore, we always have $\boldsymbol{Y}' \geq \boldsymbol{X}'$. Since $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ have the same distributions as $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, we get $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$.

3 Dependence Uncertainty

For any set of univariate marginal distribution functions $(F_1, \ldots, F_d)$, the symbol $\Gamma(F_1, \ldots, F_d)$ denotes the Fréchet class of all d-dimensional distributions having these marginals.

The main goal of this paper is to study the following set

$$\mathcal{S}_d := \{(F_1, \ldots, F_d, G_1, \ldots, G_d) : F \leq_{st} G \text{ for all } F \in \Gamma(F_1, \ldots, F_d) \text{ and all } G \in \Gamma(G_1, \ldots, G_d)\} \quad (8)$$

We start with some necessary conditions for $(F_1, \ldots, F_d, G_1, \ldots, G_d) \in \mathcal{S}_d$.

Theorem 3. *If $(F_1, \ldots, F_d, G_1, \ldots, G_d) \in \mathcal{S}_d$, then*

$$\min\{F_1(t_1), \ldots, F_d(t_d)\} \geq \max\left\{0, \sum_{i=1}^{d} G_i(t_i) - d + 1\right\} \quad \textit{for all } (t_1, \ldots, t_d) \in \mathbb{R}^d. \quad (9)$$

Proof. Assume that there exists $(t_1, \ldots, t_d)$ such that

$$\min\{F_1(t_1), \ldots, F_d(t_d)\} < \max\left\{0, \sum_{i=1}^{d} G_i(t_i) - d + 1\right\}. \quad (10)$$

Then the function

$$u(\boldsymbol{z}) = -\prod_{i=1}^{d} \mathbb{1}_{(-\infty, t_i]}(z_i) \quad (11)$$

is increasing and

$$\mathbb{E}[u(\boldsymbol{X})] = -\min\{F_1(t_1), \ldots, F_d(t_d)\} > -\max\left\{0, \sum_{i=1}^{d} G_i(t_i) - d + 1\right\} = \mathbb{E}[u(\boldsymbol{Y})],$$

if we assume that the copula of $\boldsymbol{X}$ is the upper Fréchet bound and the copula of $\boldsymbol{Y}$ is the copula $C_{\boldsymbol{t}}$ for which the lower Fréchet bound is attained in the point $\boldsymbol{t}$. Hence $\boldsymbol{X} \not\leq_{st} \boldsymbol{Y}$.

We now move to sufficient conditions, starting with a very simple but also very restrictive condition of separated supports for the marginals, which trivially yields multivariate stochastic order.

Theorem 4. *Consider the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, and assume that there exists a deterministic vector $\boldsymbol{\delta} \in \mathbb{R}^d$ such that $X_i \leq_{st} \delta_i$ and $\delta_i \leq_{st} Y_i$ for all $i = 1, \ldots, d$. Then $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$.*

Notice that this can easily be rewritten as follows:
Let $F_1, \dots, F_d, G_1, \dots, G_d$ be such that there exists $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)$ for which

$$F_i(\delta_i) = 1 - G_i(\delta_i) = 1, \quad \text{for all } i = 1, \dots, d. \tag{12}$$

Then $(F_1, \dots, F_d, G_1, \dots, G_d) \in \mathcal{S}_d$.

If we can verify this very restrictive condition on the marginals, then we can completely ignore the dependence structures of the random vectors. Notice that, in contrast to this, in Theorem 1 we had to assume equal covariance matrices even if the means are very different. This shows that we cannot expect to find much weaker sufficient conditions on the marginals that imply MSO for all dependence structures, when the supports overlap, in particular when they are unbounded.

The following result shows that in the bivariate case, if the supports of the random variables overlap only on a small region, then the relation $\leq_{st}$ holds irrespective of the dependence structure.

Theorem 5. *Let F_1, F_2, G_1, G_2 satisfy*

$$F_1(a_1) = F_2(a_2) = 1 - G_1(b_1) = 1 - G_2(b_2) = 1 \tag{13}$$

for some $a_1 > b_1, a_2 > b_2$. If

$$F_1(a_1) + F_2(a_2) \leq G_1(b_1) + G_2(b_2) - 1, \tag{14}$$

then $(F_1, F_2, G_1, G_2) \in \mathcal{S}_2$.

Proof. Define

$$A := (-\infty, b_1] \times (-\infty, b_2], \quad B := (-\infty, a_1] \times (-\infty, a_2] \setminus A, \tag{15}$$
$$D := [a_1, \infty) \times [a_2, \infty), \quad C := [b_1, \infty) \times [b_2, \infty) \setminus D. \tag{16}$$

Then $p := \mathbb{P}(\boldsymbol{X} \in A)$ fulfills

$$p \geq \mathbb{P}(X_1 \leq b_1) + \mathbb{P}(X_2 \leq b_2) - 1 \geq \mathbb{P}(Y_1 \leq a_1) + \mathbb{P}(Y_2 \leq a_2) \geq \mathbb{P}(\boldsymbol{Y} \in C) =: q.$$

For any increasing function $u : \mathbb{R}^2 \to \mathbb{R}$ we get $u(\boldsymbol{x}) \leq u(b_1, b_2) \leq u(\boldsymbol{y})$ whenever $\boldsymbol{x} \in A$ and $\boldsymbol{y} \in C$ and similarly $u(\boldsymbol{x}) \leq u(a_1, a_2) \leq u(\boldsymbol{y})$ whenever $\boldsymbol{x} \in B$ and $\boldsymbol{y} \in D$. Therefore,

$$\begin{aligned} \mathbb{E}[u(\boldsymbol{X})] &= p\mathbb{E}[u(\boldsymbol{X})|\boldsymbol{X} \in A] + (1-p)\mathbb{E}[u(\boldsymbol{X})|\boldsymbol{X} \in B] \\ &\leq pu(b_1, b_2) + (1-p)u(a_1, a_2) \leq qu(b_1, b_2) + (1-q)u(a_1, a_2) \\ &\leq q\mathbb{E}[u(\boldsymbol{Y})|\boldsymbol{Y} \in C] + (1-q)\mathbb{E}[u(\boldsymbol{Y})|\boldsymbol{Y} \in D] = \mathbb{E}[u(\boldsymbol{Y})], \end{aligned} \tag{17}$$

where the second inequality stems from $p \geq q$ and $u(a_1, a_2) \geq u(b_1, b_2)$. This concludes the proof.

The following theorem provides different sufficient conditions for dominance in the general d-dimensional case.

Theorem 6. *Let $F_1, \dots, F_d, G_1, \dots, G_d$ be such that*

$$\max\left\{0, \sum_{i=1}^{d} F_i(t) - (d-1)\right\} \;\geq\; \min\left\{1, \sum_{i=1}^{d} G_i(t)\right\} \quad \textit{for all } t \in \mathbb{R}. \tag{18}$$

Then $(F_1, \dots, F_d, G_1, \dots, G_d) \in \mathcal{S}_d$.

Proof. Let $\boldsymbol{X} = (X_1, \dots, X_d)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ be random vectors with marginals F_i and G_i, $i = 1, \dots, d$, respectively. Define

$$X^{\vee} := \max\{X_1, \dots, X_d\} \quad \text{and} \quad Y^{\wedge} := \min\{Y_1, \dots, Y_d\}. \tag{19}$$

It follows from Eqs. (11) and (12) in Rüschendorf [8] that

$$\mathbb{P}(X^{\vee} \leq t) \geq \max\left\{0, \sum_{i=1}^{d} F_i(t-) - (d-1)\right\} \quad \text{and} \quad \mathbb{P}(Y^{\wedge} \leq t) \leq \min\left\{1, \sum_{i=1}^{d} G_i(t-)\right\}, \tag{20}$$

where $F(t-) := \lim_{h \downarrow 0} F(t-h)$ is the lefthand limit of a distribution function in a point $t \in \mathbb{R}$.

If (18) holds for all $t \in \mathbb{R}$, then the same inequality holds when t is replaced by $t-$, as all involved functions are right-continuous and increasing. Thus, $X^{\vee} \leq_{st} Y^{\wedge}$, if (18) holds. Using Theorem 1 for the comonotone vectors consisting of equal components $X^{\vee}$ and $Y^{\wedge}$, respectively, we thus get

$$(X_1, \dots, X_d) \leq_{st} (X^{\vee}, \dots, X^{\vee}) \leq_{st} (Y^{\wedge}, \dots, Y^{\wedge}) \leq_{st} (Y_1, \dots, Y_d). \tag{21}$$

Transitivity of $\leq_{st}$ yields the result.

Example 3. We give an example where the condition of Theorem 5 holds, but Theorem 6 does not apply. Assume $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_2 = 0) = \mathbb{P}(Y_1 = 3) = \mathbb{P}(Y_2 = 6) = 0.9$ and $\mathbb{P}(X_1 = 2) = \mathbb{P}(X_2 = 5) = \mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_2 = 1) = 0.1$. Thus $a_1 = 2, a_2 = 5, b_1 = b_2 = 1$ and $G_1(a_1) + G_2(a_2) = 0.2 \leq 0.8 = F_1(b_1) + F_2(b_2) - 1$, which, by Theorem 5, implies $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$. However, for any $t \in (a_1, a_2)$, we get

$$\min\{1, G_1(t) + G_2(t)\} = \min\{1, 1.1\} > 0.9 = \max\{0, F_1(t) + F_2(t) - 1\}. \tag{22}$$

Therefore, condition (18) in Theorem 6 does not hold.

Condition (18) implies bounds for the supports of the random vectors and never holds for distributions with unbounded support. In particular we get the following necessary conditions for (18)

$$\sum_{i=1}^{d} F_i(t) \leq d-1 \implies G_i(t) = 0 \quad \text{for all } i = 1, \dots, d, \tag{23}$$

$$\sum_{i=1}^{d} G_i(t)) \leq 1 \implies F_i(t) = 1 \quad \text{for all } i = 1, \dots, d. \tag{24}$$

Using uniform distributions on $(0,1)$ for the marginals of $\boldsymbol{X}$ and uniform distributions on $(\varepsilon, 1+\varepsilon)$ for the marginals of $\boldsymbol{X}$, a straightforward calculation shows that we cannot improve these results in general.

Acknowledgments. We thank Ruodu Wang for some useful hints related to Theorem 5. Marco Scarsini is a member of GNAMPA-INdAM. His work is partially supported by the INdAM GNAMPA project CUP_E53C22001930001, MIUR PRIN grant 2022EKNE5K, and PRIN PNRR grant P2022XT8C8.

References

1. Kamae, T., Krengel, U., O'Brien, G.L.: Stochastic inequalities on partially ordered spaces. Ann. Probab. **5**(6), 899–912 (1977). https://doi.org/10.1214/aop/1176995659
2. Lehmann, E.L.: Ordered families of distributions. Ann. Math. Stat. **26**, 399–419 (1955). https://doi.org/10.1214/aoms/1177728487
3. Müller, A.: Stochastic ordering of multivariate normal distributions. Ann. Inst. Stat. Math. **53**(3), 567–575 (2001). https://doi.org/10.1023/A:1014629416504
4. Müller, A., Scarsini, M.: Stochastic comparison of random vectors with a common copula. Math. Oper. Res. **26**(4), 723–740 (2001). https://doi.org/10.1287/moor.26.4.723.10006
5. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. Wiley, Chichester (2002)
6. Müller, A., Scarsini, M., Tsetlin, I., Winkler, R.L.: Multivariate almost stochastic dominance: transfer characterizations and sufficient conditions under dependence uncertainty. Oper. Res. (2023). https://doi.org/10.1287/opre.2022.0596
7. Range, T.M., Østerdal, L.P.: First-order dominance: stronger characterization and a bivariate checking algorithm. Math. Program. **173**(1-2, Ser. A), 193–219 (2019). https://doi.org/10.1007/s10107-017-1213-9
8. Rüschendorf, L.: Inequalities for the expectation of Δ-monotone functions. Z. Wahrsch. Verw Gebiete **54**(3), 341–349 (1980). https://doi.org/10.1007/BF00534351
9. Rüschendorf, L.: Stochastically ordered distributions and monotonicity of the OC-function of sequential probability ratio tests. Math. Operationsforsch Stat. Ser. Stat. **12**(3), 327–338 (1981). https://doi.org/10.1080/02331888108801593
10. Scarsini, M.: Multivariate stochastic dominance with fixed dependence structure. Oper. Res. Lett. **7**(5), 237–240 (1988). https://doi.org/10.1016/0167-6377(88)90038-7
11. Scarsini, M.: Multivariate convex orderings, dependence, and stochastic equality. J. Appl. Probab. **35**(1), 93–103 (1998). https://doi.org/10.1239/jap/1032192554
12. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer, New York (2007). https://doi.org/10.1007/978-0-387-34675-5
13. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
14. Yin, C.: Stochastic orderings of multivariate elliptical distributions. J. Appl. Probab. **58**(2), 551–568 (2021). https://doi.org/10.1017/jpr.2020.104

Variable Selection for Clustering Three-Way Data

Mackenzie R. Neal(✉) and Paul D. McNicholas

McMaster University, Hamilton, ON, Canada
nealm6@mcmaster.ca

Abstract. Ample work on dimension reduction for multivariate model-based clustering has been conducted; however, to date, relatively few dimension reduction methods have been presented in the matrix-variate paradigm. Such work is, for example, useful for modelling data arising from longitudinal studies with multiple responses, multivariate repeated measures data, or image data. Similar to the multivariate paradigm, issues persist when clustering data with noisy and uninformative variables within the matrix-variate paradigm. Thus, a variable selection algorithm for the matrix-variate paradigm is presented and tested on real datasets.

Keywords: clustering · three-way data · variable selection

1 Introduction

Variable selection methods allow for the removal of uninformative variables and, in turn, improve both model fitting and model interpretability. As such, variable selection has been well-studied across statistical domains, including clustering; however, the majority of this work pertains to clustering two-way, i.e., multivariate, data via a normal mixture model. Unfortunately, many datasets cannot be modelled in such a way, e.g., because the clusters are skewed or the data are three-way. The work herein proposes extensions of a common variable selection method to the matrix-variate paradigm, allowing for modelling of longitudinal or repeated measures data, spatial data, image data, and any data with a three-way structure.

2 Finite Mixture Models

2.1 Two-Way Data

Finite mixture models arise from the assumption that a population contains subpopulations that can be modelled by a finite number of densities. A p-dimensional random vector $\mathbf{X}$ belongs to finite mixture model if the density can be written

$$f(\mathbf{x}|\boldsymbol{\vartheta}) = \sum_{g=1}^{G} \pi_g f_g(\mathbf{x}|\boldsymbol{\theta}_g),$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 317–324, 2024.
https://doi.org/10.1007/978-3-031-65993-5_39

where $\pi_g > 0$ are the mixing proportions such that $\sum_{g=1}^{G} \pi_g = 1$ and $f_g(\mathbf{x}|\boldsymbol{\theta}_g)$ are the component densities. Most commonly, these component densities are taken to be multivariate normal corresponding to the following component densities:

$$f(\mathbf{x}|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) = \sum_{g=1}^{G} \frac{\pi_g}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_g|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_g)^\top \boldsymbol{\Sigma}_g^{-1}(\mathbf{x}-\boldsymbol{\mu}_g)\right\}, \quad (1)$$

where $\boldsymbol{\mu}_g$ denotes the mean and $\boldsymbol{\Sigma}_g$ is the covariance matrix for the gth component. If we take z_{ig} to be a membership indicator where $z_{ig} = 1$ when observation i belongs to component g and 0 otherwise, then the complete-data likelihood can be written as such,

$$\mathcal{L}_c(\boldsymbol{\vartheta}) = \prod_{i=1}^{n}\prod_{g=1}^{G}[\pi_g \phi(\mathbf{x}_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)]^{z_{ig}}.$$

These component memberships, z_{ig}, are unobserved thus, the EM algorithm is used to maximize the likelihood function. This results in the replacement of z_{ig} with their corresponding expected values,

$$\hat{z}_{ig} = \frac{\hat{\pi}_g \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_g, \hat{\boldsymbol{\Sigma}}_g)}{\sum_{h=1}^{G} \hat{\pi}_h \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_h, \hat{\boldsymbol{\Sigma}}_h)}.$$

2.2 Three-Way Data

Various matrix-variate distributions are used as tools for modelling random matrices. The matrix-normal distribution can be defined as follows,

$$\phi(\mathbf{X}; \boldsymbol{\vartheta}) = (2\pi)^{pT/2}|\boldsymbol{\Sigma}|^{-T/2}|\boldsymbol{\Psi}|^{-p/2} \exp\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\boldsymbol{\Psi}^{-1}(\mathbf{X}-\mathbf{M})^\top\}, \quad (2)$$

where $\mathbf{X}_1, ..., \mathbf{X}_n$ are random matrices of dimension $r \times p$, with r representing the number of occasions and p the number of attributes. Let $\mathbf{M}$ be an $r \times p$ matrix of means, $\boldsymbol{\Psi}$ is a $r \times r$ covariance matrix, and $\boldsymbol{\Sigma}$ is a $p \times p$ covariance matrix [3]. This density can then be used as a component density in the following mixture model:

$$f(\mathbf{x}|\boldsymbol{\vartheta}) = \sum_{g=1}^{G} \pi_g \phi(\mathbf{X}; \boldsymbol{\vartheta}_g).$$

Similar to the multivariate normal, the matrix-variate normal has many nice properties, making it the most widely used matrix-variate distribution. We use this distribution to extend the `vscc` algorithm to three-way data.

3 Variable Selection

Many variable selection methods exist for normal mixtures models, and two prominent examples are `vscc` [2] and `clustvarsel` [5,9,11]. The latter is a

step-wise variable selection method wherein the Bayesian information criterion (BIC; [10]) is used to assess whether or not a given variable should be included or removed from a clustering set. The computational time for both methods is dependent on the dimensions of the dataset; however, the effect of dimensionality is much greater for the `clustvarsel` algorithm. For this reason, we extend the `vscc` algorithm to the matrix-variate paradigm. The original `vscc` algorithm is detailed in Algorithm 1. This algorithm aims to remove highly correlated variables while simultaneously identifying clearly-defined clusters. This occurs via the following criterion:

$$|\rho_{jk}| < 1 - W_j^i \quad \forall k \in V,$$

where V is the set of already selected variables and ρ_{jk} is the correlation between variable j and variable k; thus, variable j is selected if the correlation of this variable to all of the selected variables is less than the between-cluster variance of variable j, denoted by $1 - W_j$. Five potential subsets are generated by varying the exponent on W_j^i from $i = 1, \ldots, 5$ to allow for different levels of correlation in the selected set, depicted in Fig. 1. Thus, model fitting occurs a total of six times, regardless of the number of variables.

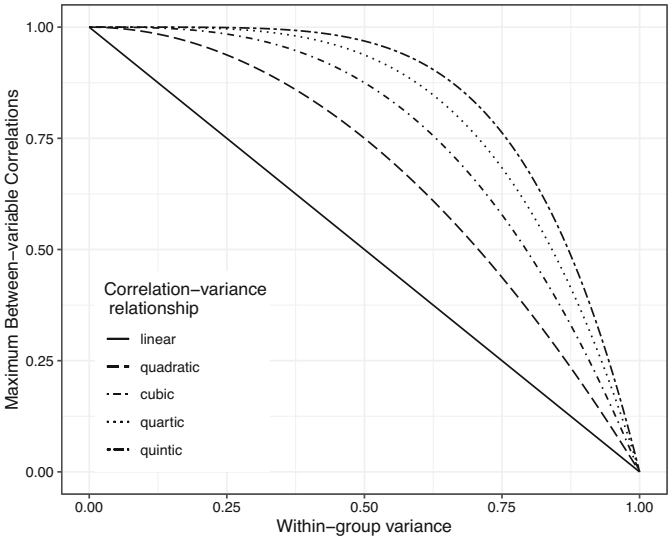


Fig. 1. Correlation-variance relationship for the `vscc` selection criteria [2]

This algorithm works well on normal clusters; however, this is a rather restrictive assumption and so Neal and McNicholas [7] have extended this algorithm for skewed components. The `vscc` package [1] for R [8] implements both the original algorithm [2] as well as the skewed extension [7]. We extend this algorithm to account for three-way data and non-normality in the following sections.

Algorithm 1 VSCC

1: Scale data and fit a mixture of multivariate normals to the scaled data.
2: Use $\hat{z}_{ig}$ as initial group memberships.
3: Calculate within-group variance for each variable,

$$\hat{W}_j = \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n}\hat{z}_{ig}(x_{ji}-\hat{\mu}_{jg})^2.$$

4: Sort $\hat{W}_j$ in ascending order.
5: **for** `i in` $1:5$ **do**
6: $\hat{W}_1$ is automatically selected into $V_{(i)}$
7: **for** `j in` $2:p$ **do**
8: **if** $|\rho_{jr}| < 1 - \hat{W}_j^i$ for all r in $V_{(i)}$ **then**
9: Variable j is placed in $V_{(i)}$
10: **else**
11: Variable j is not placed in $V_{(i)}$
12: **end if**
13: **end for**
14: **end for**
15: Perform normal mixture model-based clustering on all five variable subsets.
16: Select $V_{(i)}$ such that $n - \sum_{i=1}^{n}\max_g\{\hat{z}_{ig}\}$ is minimized.

4 Methodology

4.1 Overview

Three-way data present an interesting problem for variable selection methods pertaining to which dimension of the random matrices to reduce. One could reduce entire rows or columns of the random matrices, as would often be desirable for data arising from repeated measures experiments. Alternatively, one could remove the row-column combinations. An example wherein this may be preferable is image data, wherein each row-column combination refers to a unique pixel and thus individual pixels are removed to reduce noise in an image. We propose two extensions of the `vscc` algorithm to address each desired dimension reduction outcome for three-way data.

4.2 Method 1

The first extension of the `vscc` algorithm to three-way data will use the matrix-variate normal distribution directly. In doing so, dimension reduction will occur in only one direction, i.e., either the rows or columns of the random matrices. It is at the discretion of the user to choose which dimension to reduce; however, in cases where multiple responses of multiple measures are taken on multiple units, dimension reduction will often occur on the measures, rather than the occasions. The `vscc` algorithm for this method is given as Algorithm 2.

Results: We use the Mice Gene Expression dataset, see [4], to compare the proposed method to a model of the full dataset. This data contains 27 attributes,

Algorithm 2 VSCC Method 1

1: Scale data and fit a matrix-normal mixture model to the scaled data.
2: Use $\hat{z}_{ig}$ as initial group memberships.
3: Calculate within-group variance for each variable,

$$W_j = \frac{1}{nr}\sum_{g=1}^{G}\sum_{i=1}^{n} z_{ig}(\mathbf{X}_{jig} - \mathbf{M}_{jg})^\top(\mathbf{X}_{jig} - \mathbf{M}_{jg})$$

4: Sort $\hat{W}_j$ in ascending order.
5: **for** `i in` $1:5$ **do**
6: $\quad\hat{W}_1$ is automatically selected into $V_{(i)}$
7: $\quad$**for** `j in` $2:p$ **do**
8: $\quad\quad$**if** $|\rho_{jr}| < 1 - \hat{W}_j^i$ for all r in $V_{(i)}$ **then**
9: $\quad\quad\quad$Variable j is placed in $V_{(i)}$
10: $\quad\quad$**else**
11: $\quad\quad\quad$Variable j is not placed in $V_{(i)}$
12: $\quad\quad$**end if**
13: $\quad$**end for**
14: **end for**
15: Perform normal mixture model-based clustering on all five variable subsets.
16: Select $V_{(i)}$ such that $n - \sum_{i=1}^{n} \max_g\{\hat{z}_{ig}\}$ is minimized.

15 measurements, and 69 mice. Results are compared up to an indicator variable that indicts whether the mouse was a control mouse or a trisomic mouse, see Table 1.

Table 1. Results of clustering on the mice gene expression dataset.

Model	G	Misclassification rate	Number of variables
Matrix-normal no selection	2	0.49	27
`vscc` matrix-normal	2	0.35	7

Although both the full model and the reduced model select the correct number of groups, the misclassification rate resulting from variable selection has improved while reducing dimensions from 27×15 to 7×15.

4.3 Method 2

The second extension of the `vscc` algorithm to three-way data uses the connection to the multivariate normal [3], i.e.,

$$\mathcal{X} \sim \mathcal{N}_{r\times p}(\mathbf{M}, \mathbf{\Psi}, \mathbf{\Sigma}) \Leftrightarrow \text{vec}(\mathcal{X}) \sim \mathcal{N}_{rp}(\text{vec}(\mathbf{M}), \mathbf{\Psi} \otimes \mathbf{\Sigma}).$$

In doing so, dimension reduction will occur on the row-column combinations. This would be desirable when it is unclear whether one should reduce rows or columns. The `vscc` algorithm for this method is found in Algorithm 3.

Algorithm 3 VSCC Method 2

Vectorize data producing a $n \times rp$ observed matrix
Scale data and fit a mixture of rp-dimensional multivariate normals.
Use $\hat{z}_{ig}$ as initial group memberships.
Calculate within-group variance for each variable,

$$W_j = \frac{1}{n} \sum_{g=1}^{G} \sum_{i=1}^{n} z_{ig} (\text{vec}(\mathcal{X})_{jig} - \mu_{jg})^2$$

Sort $\hat{W}_j$ in ascending order.
for `i in` $1:5$ **do**
    $\hat{W}_1$ is automatically selected into $V_{(i)}$
    for `j in` $2:p$ **do**
        if $|\rho_{jr}| < 1 - \hat{W}_j^i$ for all r in $V_{(i)}$ **then**
            Variable j is placed in $V_{(i)}$
        else
            Variable j is not placed in $V_{(i)}$
        end if
    end for
end for
Perform normal mixture model-based clustering on all five variable subsets.
Select $V_{(i)}$ such that $n - \sum_{i=1}^{n} \max_g \{\hat{z}_{ig}\}$ is minimized.

Results: The Life Expectancy dataset [13] is used to compare all methods discussed thus far. This dataset contains 13 variables recorded for 119 countries across 16 years. A categorical variable indicating whether or not a country is considered to be under developed is used to compare clustering results up to, see Table 2. Somewhat counterintuitively, the `vscc` with the multivariate normal connection performs the best. Upon investigation, it appears that, for some years, the true clusters are more distinct than other years, likely resulting in method 2 outperforming method 1.

Table 2. Results of clustering on the life expectancy dataset.

Model	G	Misclassification rate	Number of variables
Matrix-normal no selection	4	0.62	13
`vscc` matrix-normal	6	0.62	7
Multivariate normal no selection	3	0.47	208 (13×16)
`vscc` multivariate normal	2	0.21	17

4.4 Non-normal Clusters

Both of the provided methods can be further extended for non-normal clusters, following the method proposed in Neal and McNicholas [7]. The matrix-Manly

distribution from Melnykov and Zhu [6] can be used to extend the method discussed in Sect. 4.2, and the Manly distribution from Zhu and Melnykov [14] can be used to extend the method discussed in Sect. 4.3. We provide some brief results in Table 3 from extending both methods for non-normal clusters.

Results: The Landsat Satellite dataset [12] contains 845 images, each image is photographed under four different spectral bands and is described by a 3×3 grid of pixels. Thus, resulting in 845 4×9 matrices. All proposed methods are tested on said dataset and the results are summarized in Table 3. For method 1 and its non-normal counterpart we chose to reduce the dimension pertaining to spectral bands.

Table 3. Results of clustering on the landsat satellite dataset.

Model	G	Misclassification rate	Number of variables
`vscc` matrix-normal	5	0.32	3
`vscc` matrix-Manly	3	0.27	2
`vscc` multivariate normal	4	0.42	27
`vscc` multivariate Manly	2	0.36	1

From Table 3, we see that the best performing method is the `vscc` matrix-Manly, selecting the correct number of clusters and obtaining the lowest misclassification rate. For both proposed methods, performance is improved when non-normality is considered.

5 Discussion

The literature for dimension reduction of three-way data is sparse, likely due to the unique challenge three-way data presents to dimension reduction algorithms. Nonetheless, we have shown that variable selection methods designed for two-way data can easily and effectively be extended to three-way data for both normal and non-normal clusters.

References

1. Andrews, J.L., Neal, M.R., McNicholas, P.D.: `vscc`: Variable Selection for Clustering and Classification. R Package Version 0.7 (2023). https://cran.r-project.org/package=vscc
2. Andrews, J.L., McNicholas, P.D.: Variable selection for clustering and classification. J. Classif. **31**, 136–153 (2014)
3. Gupta, A.K., Nagar, D.K.: Matrix Variate Distributions. Monographs and Surveys in Pure and Applied Mathematics Book 104. Chapman and Hall (2018)
4. Higuera, C., Gardiner, K., Cios, K.: Mice Protein Expression. UCI Machine Learning Repository (2015)

5. Maugis, C., Celeux, G., Martin-Magniette, M.L.: Variable selection for clustering with normal mixture models. Biometrics **65**(3), 701–709 (2009)
6. Melnykov, V., Zhu, X.: On model-based clustering of skewed matrix data. J. Multivar. Anal. **167**, 181–194 (2018)
7. Neal, M.R., McNicholas, P.D.: Flexible Variable Selection for Clustering and Classification. arXiv preprint arXiv:2305.16464 (2024)
8. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2023). www.R-project.org
9. Raftery, A.E., Dean, N.: Variable selection for model-based clustering. J. Am. Stat. Assoc. **101**(473), 168–178 (2006)
10. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
11. Scrucca, L., Raftery, A.E.: clustvarsel: a package implementing variable selection for normal model-based clustering in R. J. Stat. Softw. **84**(1), 1–28 (2018)
12. Srinivasan, A.: Statlog (Landsat Satellite). UCI Machine Learning Repository (1993). https://doi.org/10.24432/C55887
13. vrec99: Life Expectancy 2000–2015 (2022). https://www.kaggle.com/datasets/vrec99/life-expectancy-2000-2015
14. Zhu, X., Melnykov, V.: Manly transformation in finite mixture modeling. Comput. Stat. Data Anal. **121**, 190–208 (2018)

Functional Outlier Detection

Jeremy Oguamalam, Una Radojičić(✉), and Peter Filzmoser

Institute of Statistics and Mathematical Methods in Economics, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria
{jeremy.oguamalam,una.radojicic,peter.filzmoser}@tuwien.ac.at

Abstract. Functional data analysis is a sub-field of statistics concerned with data generated by stochastic processes $X(t) \in L^2(I), t \in I$, where I represents an interval on the real line. Unlike traditional multivariate observations, functional data are observed at individual time points throughout the interval I. This unique data structure necessitates novel approaches for analysis. The robustness of these methods is crucial to ensure reliable results, particularly in the presence of anomalies arising from measurement errors or high data variability. This study explores various frameworks designed to address these challenges, demonstrating their efficacy through simulations and real-world data analysis.

Keywords: Functional data · Functional depth · Principal component analysis · Mahalanobis distance

1 Introduction

Functional data analysis (FDA) is a statistical field that has gotten more and more recognition over the last few decades. Often the data that is observed at a number of discrete time-points $t_1, \ldots, t_p$ naturally can be viewed as a collection of functions or curves rather than just multivariate vectors $(x_{t_1}, \ldots, x_{t_p})'$. It can be assumed that the underlying data-generating process is some continuous function $t \mapsto X(t)$. Therefore, not only the observed values at the discrete time-points are relevant, but also the unobserved ones, at every other point t over some interval. Additionally, unlike for multivariate data, in FDA, the ordering of the observed data is important. Examples of functional data include e.g. daily pollution levels, yearly temperatures, accelerometer data, and stock values, to name a few. FDA also ranges over many other fields than meteorology, environmental data, medicine, and business.

The most prominent insights on FDA can be found in [19] where early theoretical developments and applications of this field are explored. A large part of their work is devoted to the representation of functional data. As common methods from multivariate statistics generally cannot be applied directly, many parametric concepts like principal component analysis (PCA), linear regression, or discriminant analysis are adapted by the authors as well. On the other hand, non-parametric methods can be found in [5]. A more recent study is the work

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 325–333, 2024.
https://doi.org/10.1007/978-3-031-65993-5_40

presented in [12]. Here, the authors target a broad mass rather than high-level experts and establish a good introductory contribution to many areas from the FDA.

Numerous statistical concepts are prone to bias in the presence of outlying observations. Consider a scenario where some observed curves deviate from the majority of the data, originating from a different process and exhibiting distinct structures. Estimating the mean or covariance in such cases becomes highly unreliable due to their sensitivity to these anomalies. Hence, robustifying statistical tools is essential. This not only mitigates bias towards outliers but also reveals crucial insights into the underlying data structure. This study aims to present current methods for detecting functional outliers. It not only adapts common multivariate tools to the functional setting but also introduces novel concepts tailored for functional data analysis.

The next section provides a brief introduction to functional data analysis, outlining key concepts. Section 3 then introduces common tools for detecting anomalies in FDA. To assess the effectiveness of these approaches, simulations are conducted and compared in Sect. 4. Section 5 concludes.

2 What Is Functional Data?

As elucidated in the introduction, functional data refers to scenarios where the data can be continuously observed and naturally leans towards modeling by curves. Here, one question the mindful reader might ask is if there even is something like "functional" data. If we observe these functions $X(t)$ just on a regular discrete grid, we could always consider them as high-dimensional multivariate observations in some way. This furthermore, suggests the questioning of the necessity for special treatment of this data structure. In [3], the author answers these statements in intuitive fashion: the idea of functional data reinforces the theoretical possibility of observing events on an infinitely fine grid over some interval I. This underlying infinite-dimensional characteristic of these mathematical objects also implies a new methodology rather than plain multivariate statistics. This further indicates that methods incorporating the special structure of functions lead to better and more reasonable analysis of such data. As we will see, some concepts cannot even be directly applied to functional data and have to be adapted consequently.

In the typical framework of FDA, the data are realizations of stochastic processes $X(t), t \in I$ where X is an element of the Hilbert space $L^2(I)$. With this notion we already inherit valuable characteristics of function spaces like the inner product $\langle \cdot, \cdot \rangle$ and a corresponding norm which are defined as

$$\langle X, Y \rangle = \int_I X(t)Y(t)dt \quad \text{and} \quad ||X|| = \sqrt{\langle X, X \rangle}$$

respectively, for two random functions $X, Y \in L^2(I)$. With this, we can already define important statistics like the covariance function

$$\begin{aligned}\gamma_X : I \times I &\to \mathbb{R} \\ (t,s) &\mapsto \mathbb{E}\left[(X(t) - \mu_X(t))(X(s) - \mu_X(s))\right]\end{aligned}$$

whereas $\mu_X(t) = \mathbb{E}[X(t)]$ is the expectation of X. This function is continuous, symmetric and positive semi-definite and furthermore satisfies

$$\int_I \int_I \gamma_X^2(t,s) ds dt < \infty.$$

For the random sample $X_1, \ldots, X_N$, the corresponding sample mean and the covariance are respectively defined as

$$\hat{\mu}(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t) \quad \text{and} \quad \hat{\gamma}_X = \frac{1}{N} \sum_{i=1}^{N} (X_i(t) - \hat{\mu}_X(t))(X_i(s) - \hat{\mu}_X(s)).$$

Additionally, the counterpart to multivariate covariance is the so-called covariance operator C_X defined as

$$(C_X f)(t) = \mathbb{E}\left[\langle X - \mu_X, f \rangle (X(t) - \mu_X(t))\right] \in L^2(I),$$

for $f \in L^2(I)$ and $t \in I$. It can be shown that this operator is a compact, self-adjoint, positive definite, integral operator with the covariance function as kernel. Therefore, for $t \in I$ it can be also represented by

$$(C_X f)(t) = \int_I \gamma_X(s,t) f(s) ds.$$

For the random sample $X_1, \ldots, X_N$, and $f \in L^2(I)$ the sample covariance estimator is defined as

$$(\hat{C}_X f) = \frac{1}{N} \sum_{i=1}^{N} \langle X_i - \hat{\mu}_X, f \rangle (X_i - \hat{\mu}_X).$$

3 Functional Outlier Detection

A frequent problem statistical tools have in common is biased results caused by outlying observations. Nowadays potential contamination in data sets has always been taken into account. In general, we say that a certain data point is an outlying observation if it differs significantly from the rest of the data set [8]. An occurrence of an outlier may be due to the large (or infinite) variability in the underlying distribution of the measured data or it may indicate some kind of an experimental or measurement error. No matter the source, outliers may cause issues in the analysis and the inference on the given data set. For more insight see e.g. [8].

In the functional setting, atypical observations can arise in many different ways. A good classification is given by the taxonomy in [9]. Isolated outliers are curves that are outlying only in a short interval of their domain. In the univariate case, this distinguishes them from the majority of the data by a small number of peaks for example. Persistent outliers, on the other hand, are functions diverging on a big part of the whole domain. These outliers are divided into shift, amplitude, and shape outliers. Shift outliers follow the main trend of the data, but their curves are shifted. Amplitude outliers also have the same shape as the majority, but their possibly significantly different scale or range is conspicuous. Shape outliers exhibit a departure from the typical or expected shape patterns of the overall dataset. The following subsections introduce methodologies based on different statistical tools that can deal with atypical functional data.

3.1 Functional Principal Component Analysis

Consider a functional random sample $X_1, \ldots, X_N$ in $L^2(I)$. Functional principal component analysis (FPCA) tries to capture the main modes of variation by decomposing them into a set of orthogonal basis functions. It starts with an optimization problem

$$\phi_1 = \underset{\phi \in L^2(I)}{\arg\max} \frac{1}{N} \sum_{i=1}^{N} \langle X_i, \phi \rangle_2^2 \quad \text{subject to} \quad ||\phi|| = 1.$$

The resulting object (ϕ_1) is called the first functional principal component (FPC) and corresponds to the direction capturing the most variability. The remaining components, ϕ_i where $i \geq 2$, correspond to directions that encapsulate progressively less variation in the data, exhibiting mutual orthogonality. The set of the first K principal components can then be used to approximate the individual curves of the random sample, i.e. $X_i \approx \sum_{i=1}^{K} c_{ik}\phi_i$ for $i = 1, \ldots, N$, where w.l.o.g $\mu_X = 0$ can be assumed. Such truncation is optimal w.r.t. mean square error, as a consequence of Kahrunen-Loève theorem; for more details see e.g. [19].

One early approach of functional outlier detection was developed in [10] where the authors proposed a method to forecast mortality and fertility rates using the functional data paradigm. Within their method, they used a weighted approach to FPCA to estimate observed time series using functional principal components. The forecast accuracy in the form of the integrated square error [19] follows a chi-squared distribution [11]. Therefore, high values of this error measurement can indicate outlyingness. This approach was e.g. used in [10], where the average squared difference between a function and its projection onto the first K robust principal components is considered as a measure of outlyingness. We consider this estimator in Sect. 4 under the name "integrated square error" (abbrev. ISE); see Fig. 2.

3.2 Functional Depth

The concept of data depth refers to the ordering of observations by some order statistic. In multivariate statistics, depth measures the "centrality" of a point within a data cloud. The higher the value, the more central the corresponding point is. In this context, the multivariate median can be defined as the point with the highest depth. Extending the notion of depth to functional data has turned out to play an important role in functional outlier detection, allowing for a natural ordering of the curves. The initial work can be found in [6], where the authors introduced the Fraimain-Muniz depth. For a random sample $X_1, \ldots, X_N \in L^2(I)$ the measure of outlyingness is based on the integral over the univariate depth D of each curve

$$FMD_i = \int_I D(X_i(t))dt, \ i = 1, \ldots, N.$$

This then leads to an intrinsic ordering by the resulting values $FMD_1, \ldots, FMD_N$. Another approach is the so-called band-depth, established in [13]. It assesses the depth of a curve within a data set relative to others by considering the proportion of curves from the dataset that lie above or below it at each point along its domain. This concept then leads to the creation of functional boxplots in [21] which are then used to detect outliers as in the multivariate setting; abbreviation "Func BP" in Fig. 2. In [4] several depth measures like the Fraiman-Muniz depth were incorporated into an iterative algorithm to flag outlying observations. These concepts are referred to as depth-based trimming and weighting, with abbrev. "DB trimming" and "DB weighting" in Fig. 2. In the work by [1], they introduced an approach that integrates information from both modified band depth [13] and half-region depth [14]. This integration results in a visualization tool designed to identify curves with distinct shapes and proposes a rule for detecting shape outliers; abbreviation "Adj. Outlierg." in Fig. 2.

3.3 Mahalanobis Distance

In the multivariate setting, many prominent methods to detect anomalies utilize the so-called Mahalanobis distance that was introduced in [16]. It calculates the distance between data points while incorporating the correlation in the data represented by the covariance. That is, data points with very high Mahalanobis distances are assumed to be far away from the center of the data cloud and are often considered outlying and examined in more detail. However, if one would use the plug-in estimator of the covariance, the effect of the outliers might be clouded by the fact that these same outliers affect the estimate of the covariance as well. Thus, a robust estimate of the covariance should be used instead. One such estimator is called the minimum covariance determinant (MCD) estimator. It is, in fact, one of the first highly robust and affine equivariant estimators of multivariate location and scatter, see e.g. [20] for more details. The main idea behind it is, given a tuning parameter h, to choose a subset of h out of N observations, for which the determinant of the sample covariance matrix is

minimal. Some of its properties are the maximal possible breakdown point of 50% for a specific choice of h, as shown in [15].

The general adaptation of the Mahalanobis distance to the functional setting is not straightforward as it would include the inverse of functional covariance. The problem is that the infinite sum over the eigenvalues of the covariance operator is finite which means that the operator is singular and the inverse operator cannot be calculated. The idea of a regularized inverse operator by [17] was utilized in [7] to introduce a functional Mahalanobis semidistance, one of the first attempts to generalize this concept to FDA. In [2] the authors took it one step further and developed a true metric based on the theory of Reproducing Kernel Hilbert Spaces; abbrev. "Mah. RKHS" in Fig. 2. For this Mahalanobis distance, which depends on a smoothing parameter α, even some distributional results as well as consistency can be derived. In [18], one of the most recent studies, the minimum regularized covariance trace (MRCT) estimator was developed. This subset-based concept incorporates the functional Mahalanobis distance in an iterative algorithm for robust covariance estimation and outlier detection.

4 Simulation

The following section should demonstrate the capability of the discussed outlier detection methods. We consider a simulation study similar to the one in [1,2] where three different outlier scenarios are considered: shift, magnitude and shape outliers. The corresponding models are

$$\begin{aligned}
\texttt{Shift:}\quad & X_1(t) = 30t(1-t)^{1.5} + \varepsilon_1(t),\ X_2(t) = 30t^{1.5}(1-t) + \varepsilon_2(t),\\
\texttt{Magnitude:}\quad & X_1(t) = 4t + \epsilon_1(t),\ X_2(t) = 4t + (-1)^u 1.8 + (0.02\pi)^{-0.5}\exp\left(\tfrac{-(t-\mu)^2}{0.02}\right) + \epsilon_2(t),\\
\texttt{Shape:}\quad & X_1(t) = 4t + \epsilon_1(t),\ X_2(t) = 4t + 2\sin(4(t+\mu)\pi) + \epsilon_2(t),
\end{aligned}$$

where X_1 and X_2 depict the regular and outlying processes, respectively. The errors $\varepsilon_1, \varepsilon_2$ and ϵ_1, ϵ_2 are independent Gaussian random processes with expectation zero and covariance functions $\gamma_{X,\varepsilon}(s,t) = 0.3\exp\left(-0.3^{-1}|s-t|\right)$ and $\gamma_{X,\epsilon}(s,t) = \exp(-|s-t|)$, respectively. Additionally we consider the random variables $u \sim \text{Bernoulli}(0.5)$ and $\mu \sim \mathcal{U}[0.25, 0.75]$. The percentage of outlying curves is set to 20% while the number of observations is $N = 200$. For each model, we consider two scenarios; each function is observed at $p = 100$ and $p = 500$ equally spaced time-points in $I = [0,1]$. A sample of curves from all three models is depicted in Fig. 1.

Each outlier model – dimension setting is repeated 50 times and the resulting classifications are collected. In Fig. 2 the overall performance of each algorithm is displayed in the form of the F-scores calculated as $TPR/(TPR + 0.5(FPR + FNR))$, where TPR (true positive rate) represents the ratio of correctly classified outliers, while the FPR (false positive rate) represents the ratio of misclassified regular observations, and FNR corresponds to the ratio of non-identified

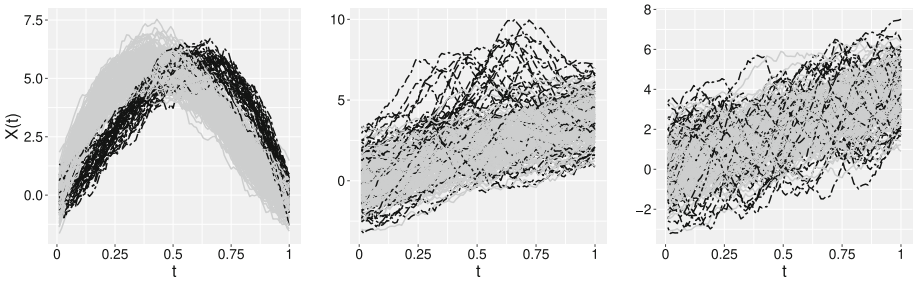


Fig. 1. Left to right, samples for shift, magnitude, and shape anomalies, respectively. Solid curves represent the main processes, while the dashed ones indicate the outliers. The contamination rate is $c = 0.2$, sample size $N = 200$, and $p = 100$ time points.

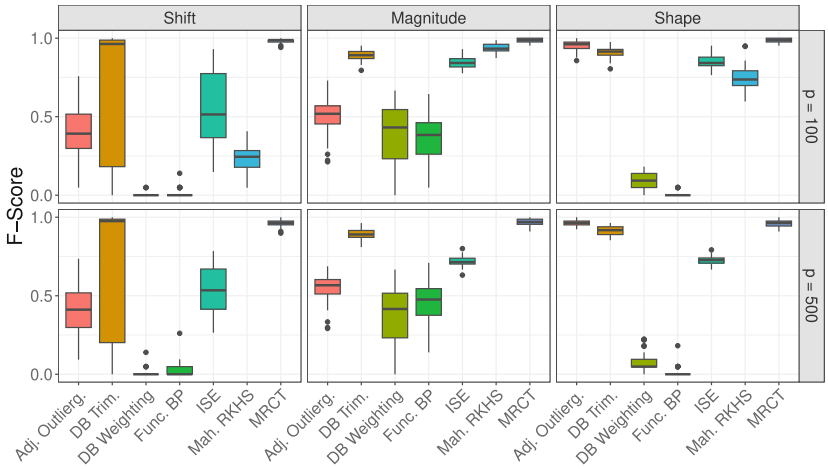


Fig. 2. Boxplots of F-score for different methods over the three models for the Gaussian process; $N = 200, p = 100, 500$, and $c = 0.2$

outliers. Due to the implementation, the Mahalanobis distance-based approach (Mah. RKHS) [2] can only be applied for $p = 100$. Depending on the specific setting, different methods seem to work better than others. Notably, the MRCT demonstrates overall noteworthy performance by effectively managing outlier detection while retaining a significant portion of regular observations.

5 Conclusion

Nowadays, finding anomalies in a data set is crucial to understand the underlying dynamics and improving the results of research questions. In this work, we presented several methods that can handle this issue effectively when the data are functional. These strategies utilize different statistical areas like functional principal component analysis, the notion of functional depth, or a generalization

of the Mahalanobis distance to handle outlying curves. A subsequent simulation study shows their good performance in many different scenarios.

Acknowledgments. This work was supported by the Austrian Science Fund (FWF), [10.55776/I5799].

References

1. Arribas-Gil, A., Romo, J.: Shape outlier detection and visualization for functional data: the outliergram. Biostatistics **15**, 603–619 (2014)
2. Berrendero, J., Bueno-Larraz, B., Cuevas, A.: On Mahalanobis distance in functional settings. J. Mach. Learn. Res. **21**(9), 288–320 (2020)
3. Cuevas, A.: A partial overview of the theory of statistics with functional data. J. Stat. Plan. Inference **147**, 1–23 (2014)
4. Febrero, M., Galeano, P., González-Manteiga, W.: Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. Environmetrics **19**, 331–345 (2008)
5. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer, Heidelberg (2006)
6. Fraiman, R., Muniz, G.: Trimmed means for functional data. TEST **10**, 419–440 (2001)
7. Galeano, P., Joseph, E., Lillo, R.E.: The Mahalanobis distance for functional data with application to classification. Technometrics **57**, 281–291 (2015)
8. Grubbs, F.E.: Procedure for detecting outlying observations in samples. Technometrics **1**, 1–21 (1969)
9. Hubert, M., Rousseeuw, P.J., Segaert, P.: Multivariate functional outlier detection. Stat. Methods Appl. **24**, 177–202 (2015)
10. Hyndman, R.J., Md. Ullah, S.: Robust forecasting of mortality and fertility rates: a functional data approach. Comput. Stat. Data Anal. **51**, 4942–4956 (2007)
11. Hyndman, R.J., Shang, H.L.: Rainbow plots, bagplots, and boxplots for functional data. J. Comput. Graph. Stat. **19**, 29–45 (2010)
12. Kokoszka, P., Reimherr, M.: Introduction to Functional Data Analysis. CRC Press, Florida (2017)
13. López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Am. Stat. Assoc. **104**, 718–734 (2009)
14. López-Pintado, S., Romo, J.: A half-region depth for functional data. Comput. Stat. Data Anal. **55**, 1679–1695 (2011)
15. Lopuhaa, H.P., Rousseeuw, P.J.: Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. Ann. Stat. **19**(1), 229–248 (1991)
16. Mahalanobis, P.C.: On the generalised distance in statistics. Proc. Indian Natl. Sci. **2**, 49–55 (1936)
17. Mas, A.: Weak convergence in the functional autoregressive model. J. Multivar. Anal. **98**, 1231–1261 (2007)
18. Oguamalam, J., Radojičić, U., Filzmoser P.: Minimum regularized covariance trace estimator and outlier detection for functional data. Technometrics (2024). https://doi.org/10.1080/00401706.2024.2336542
19. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer Series in Statistics. Springer, Heidelberg (2005)

20. Rousseeuw, P.J., Van Driessen, K.: A fast algorithm for the minimum covariance determinant estimator. Technometrics **41**, 212–223 (1999)
21. Sun, Y., Genton, M.G.: Functional boxplots. J. Comput. Graph. Stat. **20**, 316–334 (2011)

Patchwork Methods for Multivariate Quasi-Copulas

Matjaž Omladič(✉), Tomaž Košir, Damjana Kokol Bukovšek, Blaž Mojškerc, and Nik Stopar

Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
Matjaz@Omladic.net

Abstract. In this contribution, we exhibit the work presented in a series of recent papers of the authors, members of Ljubljana Copula Group. These results appear to be new since the patchwork approach has been presented for copulas so far. In the bivariate case we can find actual constructions of desired patches. It seems that the multivariate case is much more involved, so that these constructions cannot be extended to higher dimensions. Our main result is to present the conditions both on the domain of the patch and the values on its border for the desired multivariate quasi-copula to exist. This work is somewhat related to patchwork constructions of aggregation functions with given upper and/or lower margins being modular or 2-increasing.

Keywords: Copulas · Quasi-copulas · Patchwork techniques

1 Introduction

Copulas are mathematical objects that capture the dependence structure among random variables. They may be viewed either as distributions with uniform margins or as building blocks of general distributions with given margins. Since they were introduced by Sklar in 1959 [16], they have gained a lot of popularity through applications, say, in finance, insurance, and reliability theory. An important concept related to copulas are quasi-copulas which may be seen as the pointwise infima or suprema of copulas [10, Sect. 7.3]. Precise definitions will be given later. The interest in quasi-copula theory and its applications is growing fast as can be seen from a recent review by Arias-García et al. [2] (cf. also [1] and Nelsen et al. [14]).

The problem of obtaining a copula from a certain function (called *sub-copula*) defined on a suitable subset of $\mathbb{I}^n$, where $\mathbb{I} = [0, 1]$, by extending it to a copula on the whole domain $\mathbb{I}^n$, is as old as the notion of a copula itself. Indeed, even in the proof of Sklar's theorem one naturally finds the sub-copula defined on the Cartesian product of the ranges of the marginal distributions and seeks for an appropriate extension to the whole domain $\mathbb{I}^n$. With the development of the theory this question has been used in quite a number of situations. For a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 334–341, 2024.
https://doi.org/10.1007/978-3-031-65993-5_41

list of applications and a full answer to it, including a necessary and sufficient conditions on the function and its domain as well as all possible solutions of the problem, an interested reader should consult the paper by de Amo et al. [6] (cf. also [5]).

On the other hand, the analogous problem for quasi-copulas is much newer. To the best of our knowledge it has first been considered in the paper of Quesada-Molina and Sempi [15], in a very specific situation of extending a discrete (!) *sub-quasi-copula* to a quasi-copula. We exhibit here the work presented in a series of recent papers of some members of Ljubljana Copula Group [3,4]. (Nevertheless, let us point out that our work is related to [9,11,12], where patchwork techniques related to aggregation functions with given upper and/or lower margins being modular or 2-increasing have been discussed.) One can find the most general possible conditions on the function (i.e. sub-quasi-copula) and its domain, under which the function has an extension to a quasi-copula. Also, one can find all possible solutions to the problem in principle. While our conditions on domain and even on the function have some similarity to the copula case given in [6], the methods they use to solve the problem can't be applied for the case of quasi-copulas and the necessary tools have to be developed from scratch. It also turns out that the problem for the multivariate case in dimensions strictly greater than 2 is much more demanding, so that some of the methods that work in the binary case cannot be extended to higher dimensions.

2 Binary Patchwork Constructions of Quasi-Copulas

In this section, we denote by $R \subseteq \mathbb{I}^2$ a positively oriented rectangle. Consider two closed subsets δ_1 and δ_2 of $\mathbb{I}$ that both contain 0 and 1. We let $\mathcal{D} = \delta_1 \times \delta_2$. A function $Q : \mathcal{D} \to \mathbb{I}$ is called a *sub-quasi-copula* (cf. [15]) if Conditions *(i), (ii),* and *(iii)* below are fulfilled. Note that if $\delta_1 = \delta_2 = \mathbb{I}$ so that $\mathcal{D} = \mathbb{I}^2$ this is the usual definition of a quasi-copula. A copula may be seen as a quasi-copula Q such that Q-volume of every rectangle is positive.

(i) Q is grounded and 1 is the neutral element of Q;
(ii) Q is increasing in each of its arguments on $\mathcal{D}$;
(iii) Q satisfies the 1-Lipschitz condition on $\mathcal{D}$ in each of its arguments, i.e., it satisfies Condition (L) (for all $a, b, x, x', y, y' \in \mathbb{I}$):

L (a) $|Q(x', b) - Q(x, b)| \leqslant |x' - x|$
(b) $|Q(a, y') - Q(a, y)| \leqslant |y' - y|$.

Assuming that the function is given on the outside of a given rectangle $R \subseteq \mathbb{I}^2$ we want to extend it to R using an appropriate function to get a quasi-copula. The kind of procedure, sometimes called *patchwork*, was first introduced by De Baets and De Meyer [7] for a class of functions containing copulas (cf. also Durante et al. [8], Durante and Sempi [10, p. 115], and the references given there). From what is said there it is understood that it is appropriate for functions such as copulas that allow linear rescaling on the sides of a rectangle (cf. the

remark at the end of [7, Thm. 1]). Clearly, quasi-copulas do not satisfy this condition in general and the same source gives for them only necessary and sufficient conditions on a patch to fit into a quasi-copula (cf. [7, Prop. 6]). Using the characterization considered here we indeed notice that we need non-linear scaling.

Let the values on the boundary ∂R of the rectangle $R = [a, b] \times [c, d] \subseteq \mathbb{I}^2$ be given via univariate functions: $F(x)$ respectively $F'(x)$ for $x \in [a, b]$ on the south side respectively the north side, and $G(y)$ respectively $G'(y)$ for $y \in [c, d]$ on the west side respectively the east side. Moreover, assume that Q is a quasi-copula such that its values on ∂R coincide with the values given above.

Conditions PB (*Properties of the boundary functions*). The functions F, F', G, and G' have to satisfy the following conditions.

(i) The values of these functions are uniquely determined at the vertices of R meaning that $\alpha := F(a) = G(c)$, $\beta := F(b) = G'(c)$, $\gamma := F'(b) = G'(d)$, and $\delta := F'(a) = G(d)$.
(ii) These functions are increasing and 1-Lipschitz.
(iii) For all $x \in [a, b]$ and $y \in [c, d]$ we have $0 \leqslant F'(x) - F(x) \leqslant d - c$ and $0 \leqslant G'(y) - G(y) \leqslant b - a$.

Moreover, we will assume that the differences $F' - F$ and $G' - G$ are both monotone and non-constant on their respective domains $[a, b]$ and $[c, d]$, so that the "signed lengths" of the ranges are equal to

$$(\gamma - \beta) - (\delta - \alpha) = V = (\gamma - \delta) - (\beta - \alpha),$$

where V is the Q-volume of R. This shows, in particular, that either these two differences are both increasing or they are both decreasing depending on whether $V > 0$ or $V < 0$. Clearly, under our assumption V is nonzero since the two differences are assumed non-constant. If $V > 0$, then $(F' - F - \delta + \alpha)/V : [a, b] \to [0, 1]$ and $(G' - G - \beta + \alpha)/V : [c, d] \to [0, 1]$ are surjective increasing mappings. If $V < 0$ then we deduce exactly the same with a careful reexamination of this reasoning. Now, choose an arbitrary quasi-copula D and interpolate it into the patch. Actually, for $(x, y) \in R$ we let

$$Q(x, y) = F(x) + G(y) - \alpha + V\widetilde{D}(x, y), \tag{1}$$

where

$$\widetilde{D}(x, y) = D\left(\frac{F'(x) - F(x) - (\delta - \alpha)}{V}, \frac{G'(y) - G(y) - (\beta - \alpha)}{V}\right).$$

It is clear from the above that D at given arguments x and y is well-defined. It turns out that the function so defined really fits the patch, and that it is increasing and 1-Lipschitz in each argument.

Theorem 1 (Elementary patch). *[4, Thm. 1] Let Q be a quasi-copula and let $R \subseteq \mathbb{I}^2$ be a rectangle. Moreover, let F, F', G, and G' satisfy Conditions* **PB**, *and let the differences $F' - F$ and $G - G'$ be monotone. Then Eq. (1) gives a possible patch that can replace the given patch of Q yielding again a quasi-copula.*

It is obvious that not every extension of a sub-quasi-copula can be obtained in this way. One can have a variety of sign-changes within a patch. The situation is even more complicated for dimensions $n > 2$. We can show by counterexamples that no constructions of the kind above are possible in this case.

3 The Multivariate Case—Preliminaries

The definition of a *sub-quasi-copula* and all the accompanied notions introduced in Sect. 2 can be extended easily to the multivariate case with $n \geqslant 2$. In the *patchwork* approach we first assume that the function is given on the outside of a given rectangle $R = \prod_{i=1}^{n}[a_i, b_i] \subseteq \mathbb{I}^n$, and we want to extend it to R using an appropriate function to obtain a quasi-copula. Let a function be given on ∂R, i.e., on all $2n$ $(n-1)$-faces of R called simply faces. They are given in pairs: For each $k \in [n] = \{1, 2, \ldots, n\}$ we have the face defined by $x_k = a_k$ respectively $x_k = b_k$, which we call the lower and upper k-th face and denote them by R_k and R'_k. We assume (inductively) that the values are given via $(n-1)$-variate functions: $F_k(\mathbf{x})$ respectively $F'_k(\mathbf{x}')$ for $\mathbf{x} \in R_k$ respectively $\mathbf{x}' \in R'_k$ for all $k \in [n]$. By abuse of notation, we will identify R_k with R'_k by viewing an $\mathbf{x}$ from the first one as corresponding to an $\mathbf{x}'$ from the latter one if they are equal in all coordinates except the k-th one. We rewrite the n-variate Conditions **PB** on a certain function F defined on the lower faces and by F' defined on the upper ones, with additional subscript identifying the actual face.

Conditions PB (*Properties of the boundary functions*). The functions F_k, F'_k for $k \in [n]$ have to satisfy the following conditions.

(i) The values of these functions are uniquely determined at the $(n-2)$-faces of R at which faces R_k and R'_k intersect.[1]
(ii) These functions are increasing and 1-Lipschitz in each variable.
(iii) For all $k \in [n]$ and all $\mathbf{x} \in R_k$, we have $0 \leqslant F'_k(\mathbf{x}) - F_k(\mathbf{x}) \leqslant b_k - a_k$.

By abuse of notation, we first introduce formally a function F of n variables from ∂R such that its value at $x_k = a_k$ equals F_k, and such that its value at $x_k = b_k$ equals F'_k for $k \in [n]$. Since we will always need this function only at points that belong to the boundary ∂R, there will be no ambiguity in this notation even if such function defined on all R does not exist. For an arbitrary $\mathbf{x} \in R$ and $z \in \{-1, 0, 1\}^{[n]}$ let $\mathbf{x}^z$ be the n-tuple whose k-th coordinate is equal to

$$(\mathbf{x}^z)_k = \begin{cases} b_k, \text{ if } z(k) = 1; \\ x_k, \text{ if } z(k) = 0; \\ a_k, \text{ if } z(k) = -1; \end{cases}$$

for $k \in [n]$. Because of the requirement above that F be defined on ∂R only, the value $F(\mathbf{x}^z)$ will always be defined as soon as we insist that $z \neq 0$.

[1] This means that for every pair $j, k \in [n], j \neq k$, we have four equalities of the type, say, function $F_j(\mathbf{x})$ at $\mathbf{x} \in R_j$ with $x_k = a_k$ equals $F_k(\mathbf{x})$ at $\mathbf{x} \in R_k$ with $x_j = a_j$.

Proposition 1 (Local patch bounds). *[4, Prop. 4] Fix a choice of function F satisfying the boundary conditions* **PB** *(i)–(iii). Then the set $\mathcal{S}$ of patches having F as its boundaries, and being increasing and 1-Lipschitz in each argument, is nonempty. The local bounds of $\mathcal{S}$ on R are equal to*

$$\overline{\mathcal{S}} = \min_{k\in[n]}\{F(\mathbf{x}^{z'_k}), F(\mathbf{x}^{z_k}) + x_k - a_k\}$$

$$\underline{\mathcal{S}} = \max_{k\in[n]}\{F(\mathbf{x}^{z_k}), F(\mathbf{x}^{z'_k}) + x_k - b_k\}.$$

4 The Multivariate Case—Extension

Although the proof of our main result promised in Sect. 1 is quite involved we will try to exhibit here some of the main ideas of the proof. Suppose that $Q^* : \mathcal{D} \to \mathbb{I}$ is a sub-quasi-copula, defined on $\mathcal{D} = \delta_1 \times \delta_2 \times \cdots \times \delta_n$, where the closed sets $\{0,1\} \subseteq \delta_i \subseteq \mathbb{I}$ are given for $i \in [n]$. Also, for $i \in [n]$ let us introduce:

(i) Write the sets δ_i as a union of singletons and closed intervals. Consider only maximal possible intervals contained in δ_i, and observe that any two of the kind are disjoint. Denote by $\mathcal{I}_i = \{I^i_j\}_{j\in J_i}$ the (necessarily countable) set of all such nontrivial intervals.
(ii) Write the complement $\mathbb{I} \setminus \delta_i$ as a union of countably many disjoint open intervals. We introduce the family $\mathcal{O}_i = \{O^i_j\}_{j\in K_i}$ of all the closures of these intervals.
(iii) Introduce $\mathcal{T}_i = \mathcal{I}_i \cup \mathcal{O}_i$ and $L_i = J_i \sqcup K_i$.

Observe that all the index sets J_i, K_i, and L_i are countable. Also, any pair of the closed intervals in the family $\mathcal{T}_i$ is clearly either disjoint or intersects at a common endpoint. By abuse of notation, we will identify a family of intervals by the union of these intervals. Observe that $\mathcal{T}_i$ is dense in $\mathbb{I}$. Indeed, assume, if possible, that $\mathbb{I}$ contains an open interval I having an empty intersection with $\mathcal{T}_i$. Therefore, I has an empty intersection with $\mathcal{O}_i$, so that it is contained in δ_i. Consequently, there is a closed interval containing I and being a member of $\mathcal{I}_i$, contradicting the above. Note that the complement of $\mathcal{T}_i$ in $\mathbb{I}$ can be a countable or even uncountable Cantor like set with positive Lebesgue measure in [0,1] (cf. [6,13]).

Let us denote by $\mathcal{T}$ the set of all rectangles $R_{\mathbf{s}} = \prod_{i=1}^{n}[a_{s_i}, b_{s_i}]$ indexed by n-tuples of indices $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ with $s_i \in L_i$ for $i \in [n]$ and denote the set of these n-tuples by L. Observe that necessarily $[a_{s_i}, b_{s_i}]$ is either a member of $\mathcal{I}_i$ or a member of $\mathcal{O}_i$, and that in either case $a_{s_i} \neq b_{s_i}$, so that a rectangle $R_{\mathbf{s}}$ for $\mathbf{s} \in L$ is always nondegenerate. It follows easily from our definitions that the family of rectangles $R_{\mathbf{s}}$ for $s \in L$ is countable and dense in $\mathbb{I}^n$. Following our notation of Sect. 2, we introduce a function $F_{\mathbf{s}}$ to be defined on the boundary of the rectangle $R_{\mathbf{s}}$ for every $\mathbf{s} \in L$. It is almost clear from the above that $F_{\mathbf{s}}$ is defined at least at the vertices of $R_{\mathbf{s}}$, i.e., at points $\mathbf{x}^z$ where vector z has no

zero entries. Let us look into some details about where exactly this function is defined . In the defining product $R_{\mathbf{s}} = \prod_{i=1}^{n}[a_{s_i}, b_{s_i}]$ some of the segments (say, k of them, $0 \leqslant k \leqslant n$) belong to $\mathcal{I}$. and some to $\mathcal{O}$.. Let us rewrite it so that the former ones are counted with starting indices, continuing with counting the latter ones, after an appropriate permutation of indices:

$$\widetilde{R}^k_{\mathbf{s}} = \prod_{i=1}^{n}[a_{s_i}, b_{s_i}], \text{ where } [a_{s_i}, b_{s_i}] \in \mathcal{I}_i, 1 \leqslant i \leqslant k, \text{ and } [a_{s_i}, b_{s_i}] \in \mathcal{O}_i, k < i \leqslant n.$$

Consequently, the definition set of the function $F_{\mathbf{s}}$ contains exactly the members of $\widetilde{R}^k_{\mathbf{s}}$ of the form $\mathbf{x}^z$ with z_i nonzero whenever $k < i \leqslant n$. Clearly, the definition set of $F_{\mathbf{s}}$ on the original rectangle $R_{\mathbf{s}}$ can be obtained from this one using the reverse permutation of indices on this result. We will denote it by D_F. It is our aim to find a possible quasi-copula Q solving our problem using three main ingredients: (1) use induction on dimension k for which Q^* is defined on a given $R_{\mathbf{s}}$ for $\mathbf{s} \in L$; (2) extend these solutions on all of $\mathcal{T}$ using the fact that L is countable, and (3) extend this solution by continuity to $\mathbb{I}^n$.

Here is the key result of our investigations. Recall that the main conditions on a quasi-copula (after taking care of the boundary conditions) are that it be monotone and 1-Lipschitz in each argument. So, it turns out that the following conditions are necessary for a sub-quasi-copula to be further extendable to a quasi-copula. We will write them in notation adjusted to $\widetilde{R}^k_{\mathbf{s}}$, where k is the number of segments that belong to $\mathcal{I}$. (Observe that our induction is to be performed on k.)

(A) In all variables s_i for all $1 \leqslant i \leqslant k$ the function $F_{\mathbf{s}}$ must satisfy Conditions **PB**.

(B) Let for any $j \in [n]$ and some $\mathbf{s}, \mathbf{s}' \in L$ such that $s_i = s'_i$ for $i \in [n], i \neq j$, and $b_{s_j} \leqslant a_{s'_j}$. Then we have $0 \leqslant F_{\mathbf{s}'}(\mathbf{x}^{z_j}) - F_{\mathbf{s}}(\mathbf{x}^{z'_j}) \leqslant a_{s'_j} - b_{s_j}$ for all $\mathbf{x} \in D_F \subseteq R_{\mathbf{s}}$,[2] where z_j, z'_j are defined as just before Prop. 1.

Condition (B) implies, in particular, that if $b_{s_j} = a_{s'_j}$, then $F_{\mathbf{s}'}(\mathbf{x}^{z_j}) = F_{\mathbf{s}}(\mathbf{x}^{z'_j})$ for all $\mathbf{x} \in R_{\mathbf{s}}$. In other words, if two rectangles have all but one indeces in common, so that they intersect in an $(n-1)$-face, their functions F have the same value on this intersection.

Theorem 2. *[3, Thm. 7] Fix a sub-quasi-copula Q^* and a rectangle $R_{\mathbf{s}}, \mathbf{s} \in L$.*

(a) Assume that for the rectangle considered the value of the boundary function $F_{\mathbf{s}}$ has been extended to all faces of dimension up to a certain index k and choose a face of dimension $k+1$. Then, the boundary function $F_{\mathbf{s}}$ can be extended to this face satisfying Conditions (A) and (B).

[2] Observe that this is the same as requiring $\mathbf{x} \in R_{\mathbf{s}'}$. Consequently, D_F are the same on both rectangles.

(b) *The function F can be extended to a patch on the whole rectangle considered satisfying Condition (A) and (B). The collection of these patches (when going through all the rectangles in a certain order and taking into account the definitions of functions F on previous rectangles) defines a quasi-copula extension of the given sub-quasi-copula, after extending it to the closure by continuity.*

(c) *Every quasi-copula extending the given sub-quasi-copula can be obtained in this way.*

Once we go through all possible rectangles $R_{\mathbf{s}}$ ordered in all possible ways and choose all possible patches on the way, we get all possible solutions of the given problem.

Acknowledgments. The authors are thankful for many fruitful discussions on some themes of this paper to the following colleagues: David Dolžan, Peter Erich Klement, and Susanne Saminger-Platz. The authors Tomaž Košir and Matjaž Omladič acknowledge financial support from the ARIS (Slovenian Research and Innovation Agency, research core funding No. P1-0448).

References

1. Arias-García, J.J., Mesiar, R., De Baets, B.: The unwalked path between quasi-copulas and copulas: stepping stones in higher dimensions. Intern. J. Appr. Reason. **80**, 89–99 (2017)
2. Arias-García, J.J., Mesiar, R., De Baets, B.: A Hitchhiker's guide to quasi-copulas. Fuzzy Sets Syst. **393**, 1–28 (2020)
3. Bukovšek, D.K., Košir, T., Mojškerc, B., Omladič, M.: Extending Multivariate Sub-Quasi-Copulas. arXiv:2309.15824 (2023)
4. Bukovšek, D.K., Košir, T., Omladič, M., Stopar, N.: Extending sub-quasi-copulas. J. Math. Anal. Appl. **500**, 125099 (2021)
5. de Amo, E., Díaz Camillo, M., Fernández-Sánchez, J.: Characterization of all copulas associated with non-continuous random variables. Fuzzy Sets Syst. **191**, 103–112 (2012)
6. de Amo, E., Díaz Camillo, M., Durante, F., Fernández-Sánchez, J.: Extensions of subcopulas. J. Math. Anal. Appl. **452**, 1–15 (2017)
7. De Baets, B., De Meyer, H.: Orthogonal grid constructions of copulas. IEEE Trans. Fuzzy Syst. **15**, 1053–1062 (2007)
8. Durante, F., Fernández-Sánchez, J., Quesada-Molina, J.J., Úbeda-Flores, M.: Convergence results for patchwork copulas. Eur. J. Oper. Res. **247**, 525–531 (2015)
9. Durante, F., Fernández-Sánchez, J., Sempi, C.: Multivariate patchwork copulas: a unified approach with applications to partial comonotonicity. Insur. Math. Econ. **53**, 897–905 (2013)
10. Durante, F., Sempi, C.: Principles of Copula Theory. CRC/Chapman & Hall, Boca Raton (2015)
11. Durante, F., Saminger-Platz, S., Sarkoci, P.: On representations of 2-increasing binary aggregation functions. Inf. Sci. **178**(23), 4534–4541 (2008)
12. Durante, F., Saminger-Platz, S., Sarkoci, P.: Rectangular patchwork for bivariate copulas and tail dependence. Commun. Stat. Theory Methods **38**(15), 2515–2527 (2009)

13. Gelbaum, B.R., Olmsted, J.M.H.: Counterexamples in Analysis. Dover Publications (2003)
14. Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M.: Some new properties of quasi-copulas. In: Cuadras, C.M., Fortiana, J., Rodríguez-Lallena, J.A. (eds.) Distributions with Given Marginals and Statistical Modelling, pp. 187–194. Kluwer Academic Publishers, Dordrecht (2002)
15. Quesada Molina, J.J., Sempi, C.: Discrete quasi-copulas. Insur. Math. Econ. **37**, 27–41 (2005)
16. Sklar, A.: Fonctions de répartition à n dimensions et leurs Marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)

Optimal Change-Point Sequential Detection in Autoregressive Time Series

Serguei Pergamenchtchikov[1,2] and Roman Tenzin[1(✉)]

[1] Univ Rouen Normandie, CNRS, Normandie University, LMRS UMR 6085, 76000 Rouen, France
mr.roman.tenzin@gmail.com

[2] International Laboratory SSP-QF, Tomsk State University, Tomsk, Russia

Abstract. In this study, we consider a sequential detection problem in Bayesian setting for autoregressive times series based on a bounded number of observations under the condition that the post-change parameters are unknown. To this end we propose a new truncated sequential detection method through the theory developed in [5] for the statistical model with known post-change distributions. Based on the developed method, the quickest detection algorithm is proposed, that is, optimal in terms of the minimum mean time delay with the probability of a false alarm limited by some fixed known threshold.

Keywords: autoregressive times series · sequential detection · optimal stopping · Bayesian setting

1 Model

In this paper we consider autoregressive varying coefficients time series of order $p \geq 1$ defined as

$$X_n = \vartheta_{1,n} X_{n-1} + \ldots + \vartheta_{p,n} X_{n-p} + \xi_n\,, \quad 1 \leq n \leq N, \tag{1}$$

where the initial values $X_0 = \mathbf{u}_1, \ldots, X_{1-p} = \mathbf{u}_p$ are fixed nonrandom constants, $\vartheta_{i,n} = \vartheta_i^* \mathbf{1}_{\{n \leqslant \nu\}} + \vartheta_i \mathbf{1}_{\{n > \nu\}}$, the change-point ν is an unknown integer number from $\{0, 1, \ldots, N\}$ and $(\xi_n)_{1 \leq n \leq N}$ are i.i.d. Gaussian random variables with the parameters $(0, 1)$. Moreover, we assume that the parameter vector $\theta^* = (\vartheta_1^*, \ldots, \vartheta_p^*)$ is known and the post change parameter vector $\theta = (\vartheta_1, \ldots, \vartheta_p)$ is unknown and belongs to the finite set of possible values $\Theta = \{\theta_1, \ldots, \theta_m\} \subset \mathbb{R}^p$ and $\theta^* \notin \Theta$. Such models are very popular in the times series analysis (see, for example, in [1,7,8] and the references therein). The problem is to detect the moment ν as soon as possible for any fixed $N \geq 1$. It should be emphasized that we are considering the case of small values such as $N = 5, N = 10$ or $N = 30$, which are common in many practical problems such as epidemic [5] or signal detection in the case, where the signal appears during a short time

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 342–348, 2024.
https://doi.org/10.1007/978-3-031-65993-5_42

interval even for $N = 1$ (see, for example, in [2,4] and the references therein). To this end, we use the approach proposed in [5] for the case when the parameter vector θ is known, and with accordance to which this problem is studied in the Bayesian setting with the uniform prior distribution for ν, i.e. we assume that ν is a random variable with values in the set $\{0, 1, 2, \ldots, N\}$, independent of observations $(X_n)_{1 \le n \le N}$ and for any $0 \le n \le N$ the prior probability

$$\mathbf{P}\big(\nu = n\big) = \frac{1}{N+1}\,. \tag{2}$$

In this paper we study this problem in the sequential analysis setting, i.e. we need to construct a sequential procedure to detect the change-point as soon as possible. In this case the sequential procedure is defined through a stopping time τ, which is a random variable with values in the set $\{0, 1, \ldots, N\}$ such that for any $0 \le n \le N$ the event $\{\tau = n\}$ belongs to the σ - field generated by the observations $(X_1, \ldots, X_n)$, i.e $\{\tau = n\} \in \sigma\{X_1, \ldots, X_n\}$. Now we need to define the detection risk for this problem. To this end, note that the observations (1) form a Markov chain with memory p, i.e. for any Borel set $\Gamma \subset \mathbb{R}$ and for any $1 \le n \le N$

$$\mathbf{P}\big(X_n \in \Gamma \,|\, X_1, \ldots, X_{n-1}\big) = \mathbf{P}\big(X_n \in \Gamma \,|\, Y_{n-1}\big)\,,$$

where $Y_n = \big(X_n, \ldots, X_{n-p+1}\big)'$ and the prime $'$ means the transposition. Therefore, the corresponding $X_n \,|\, X_1, \ldots, X_{n-1}$ conditional densities before and after ν are defined as

$$f^*(x \,|\, x_1, \ldots, x_{n-1}) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-(\theta^*)'\, y_{n-1})^2}{2}}$$

and

$$f^{\theta}(x \,|\, x_1, \ldots, x_{n-1}) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-(\theta)'\, y_{n-1})^2}{2}}\,,$$

where $y_n = (x_n, \ldots, x_{n-p+1})'$ with $x_i = \mathbf{u}_{1-i}$ for $-p+1 \le i \le 0$. Now for this problem one can use the general change - point model proposed in [3,6,7]. To this end, it should be noted here that the density of the random vector $(X_1, \ldots, X_n)$ in $\mathbb{R}^n$ for any $1 \le n \le \nu$ is given as

$$\mathbf{p}^*(x_1, \ldots, x_n) = \frac{1}{\big(2\pi\big)^{n/2}}\, e^{-\frac{1}{2}\,\sum_{j=1}^{n}(x_j-(\theta^*)'\, y_{j-1})^2}$$

and for $\nu < n$ this density is defined as

$$\mathbf{p}_{\nu}^{\theta}(x_1, \ldots, x_n) = \frac{1}{\big(2\pi\big)^{n/2}}\, e^{-\frac{1}{2}\,\sum_{j=1}^{\nu}(x_j-(\theta^*)'\, y_{j-1})^2}\, e^{-\frac{1}{2}\,\sum_{j=\nu+1}^{n}(x_j-(\theta)'\, y_{j-1})^2}\,.$$

Then, for any Borel set $\Gamma \subseteq \mathbb{R}^n$ we define the corresponding probabilities as

$$\mathbf{P}_n^*(\Gamma) = \int_{\Gamma} \mathbf{p}^*(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

and

$$\mathbf{P}^{\theta}_{\nu,n}(\Gamma) = \int_{\Gamma} \mathbf{p}^{\theta}_{\nu}(x_1,\ldots,x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_n\,.$$

In the sequel we set $\mathbf{P}^{*} = \mathbf{P}^{*}_{N}$ and $\mathbf{P}^{\theta}_{\nu} = \mathbf{P}^{\theta}_{\nu,N}$. Moreover, to take into account the prior distribution (2) we need the mixture probability measure, which for any $I \subseteq \{0,1,\ldots,N\}$ and any Borel set $\Gamma \subseteq \mathbb{R}^N$ is defined as

$$\widetilde{\mathbf{P}}(I\times\Gamma) = \frac{1}{m\,(N+1)}\sum_{j=1}^{m}\sum_{i\in I}\mathbf{P}^{\theta_j}_{i}(\Gamma)\,. \tag{3}$$

To formulate the detection problem we denote by $\mathcal{M}_{\alpha}$ the set of all stopping times for which *Probability of a False Alarm* (PFA) is less than some fixed threshold $0 \le \alpha \le 1$, i.e. $\widetilde{\mathbf{P}}(\tau < \nu) \le \alpha$. For any $\tau \in \mathcal{M}_{\alpha}$ we set the average delay risk as

$$\mathcal{R}(\tau) = \widetilde{\mathbf{E}}(\tau-\nu)_{+}\,,$$

where $\widetilde{\mathbf{E}}$ is the expectation with respect to the mixture measure (3) and $(x)_{+}$ is the positive part of x, i.e. $(x)_{+} = \max(x,0)$. The detection problem is to minimise this risk, i.e.

$$\inf_{\tau\in\mathcal{M}_{\alpha}}\mathcal{R}(\tau)\,. \tag{4}$$

It should be noted that such problems are considered only in asymptotic setting, i.e. when $\alpha \to 0$ and $N = +\infty$ (see, for example, in [7,9] and the references therein). We study this problem on the basis of the Lagrange multiplier method, i.e. we consider the problem

$$\inf_{\tau\in\mathcal{M}}\Big(\lambda\,\widetilde{\mathbf{E}}(\tau-\nu)_{+} + \widetilde{\mathbf{P}}(\tau<\nu)\Big)\,, \tag{5}$$

where now $\mathcal{M}$ is the class of all stopping times (without constraints) and the $\lambda > 0$ is a Lagrange multiplier. To study this problem similarly to method developed in [5] we use the weighted Shiryaev–Roberts statistics

$$R^{\theta}_{n} = \frac{1}{N+1}\sum_{k=0}^{n} h^{\theta}_{k,n} \tag{6}$$

and $h^{\theta}_{k,n}$ are the Radon-Nikodym densities of $\mathbf{P}^{\theta}_{k,n}$ with respect to $\mathbf{P}^{*}_{n}$, i.e.

$$\begin{aligned} h^{\theta}_{k,n} &= \frac{\mathrm{d}\mathbf{P}^{\theta}_{k,n}}{\mathrm{d}\mathbf{P}^{*}_{n}} = \frac{\mathbf{p}^{\theta}_{k}(X_1,\ldots,X_n)}{\mathbf{p}^{*}(X_1,\ldots,X_n)} \\ &= e^{\frac{1}{2}\left(\sum_{j=k+1}^{n}(X_j-(\theta^{*})'\,Y_{j-1})^2-\sum_{j=k+1}^{n}(X_j-\theta'\,Y_{j-1})^2\right)}\,. \end{aligned}$$

Note that unlikely to case studied in [5] (when θ is known) we can't calculate directly the statistics (6). For this reason, we will replace R^{θ}_{n} with its averaging over the parameter, i.e.

$$\overline{R}_{n} = \frac{1}{m}\sum_{i=1}^{m} R^{\theta_i}_{n}\,.$$

Similarly to [5] we can write that for any $\tau \in \mathcal{M}$

$$\lambda \widetilde{\mathbf{E}}\,(\tau-\nu)_+ + \widetilde{\mathbf{P}}\,(\tau<\nu) = 1 - \mathbf{E}^*\,G_\tau\,,$$

where $G_n = \overline{R}_n - \lambda \sum_{j=0}^{n-1} \overline{R}_j$ and $\mathbf{E}^*$ is expectation over the probability $\mathbf{P}^*$. Thus, the problem (5) could be represented like

$$\max_{\tau\in\mathcal{M}} \mathbf{E}^*\,G_\tau\,. \tag{7}$$

Now, to apply the dynamic programming method to this problem one needs to represent the function G as a function of a homogeneous Markov chain. To do this, first note that the sequence $(Y_n)_{0\le n\le N}$ for $\nu = N$ can be represented in $\mathbb{R}^p$ as

$$Y_n = A\,Y_{n-1} + \xi_n\,\mathbf{e}_1\,, \tag{8}$$

where

$$A = \begin{pmatrix} \theta_1^* & \cdots & \theta_{p-1}^* & \theta_p^* \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^p\,.$$

That means that under the probability $\mathbf{P}^*$ the process (8) is a homogeneous Markov chain with the values in $\mathbb{R}^p$. Moreover, from (6) we get immediately that for any $\theta \in \Theta = \{\theta_1, \ldots, \theta_m\}$

$$R_n^\theta = \eta^\theta(X_n, Y_{n-1})\,R_{n-1}^\theta + \frac{1}{N+1}\,,$$

where for $x \in \mathbb{R}$ and $y \in \mathbb{R}^p$ the function η^θ is defined as

$$\eta^\theta(x,y) = e^{\frac{1}{2}\left((x-y'\theta^*)^2-(x-y'\theta)^2\right)}\,.$$

Therefore, denoting by $V_n = (R_n^{\theta_1}, \ldots, R_n^{\theta_m})'$, we obtain that

$$V_n = \mathbf{D}(X_n, Y_{n-1})\,V_{n-1} + \mathbf{e}_2\,,$$

where the $m \times m$ matrix

$$\mathbf{D}(x,y) = \mathrm{diag}(\eta^{\theta_1}(x,y), \ldots, \eta^{\theta_m}(x,y)) \quad \text{and} \quad \mathbf{e}_2 = \frac{1}{N+1}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^m\,.$$

This implies that the process $Z_n = (V_n', Y_n')$ is a homogeneous Markov chain with value in $\mathbb{R}^{m+p}$ and the problem (7) can be represented as a stopping time problems for the function G of the chain $(Z_n)_{0\le n\le N}$. Now we can apply to the problem (7) the optimal stopping methods developed through dynamic programming approach in [5]. To do this, we define the transition operator for this

process, i.e. for any bounded measurable $\mathbb{R}^{m+p} \to \mathbb{R}$ function h and for any vector $z = (v, y) \in \mathbb{R}^{m+p}$ we set

$$T(h)(z) = \mathbf{E}^* \left(h(Z_1) \,|\, Z_0 = z \right).$$

Note that in our case this operator can be represented as

$$T(h)(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} h\big(\widetilde{\mathbf{D}}(u, y)\, v + \mathbf{e}_2, A\, y + u\, \mathbf{e}_1\big)\, e^{-\frac{u^2}{2}}\, \mathrm{d}u\,,$$

where $\widetilde{\mathbf{D}}(u, y) = \mathbf{D}(y'\theta^* + u, y)$. Now, we need to enter the operator that defines the Snell envelope for the process $(G_n)_{0 \le n \le N}$, i.e. we set

$$\mathbf{Q}_\lambda(h)(z) = \max\Big(h(z),\, \mathbf{T}\big(h\big)(z) - \lambda\, g(z) \Big)\,,$$

where for any $z = (v, y) = (v_1, \ldots, v_m, y_1, \ldots, y_p)$ from $\mathbb{R}^{m+p}$

$$g(z) = \frac{1}{m} \sum_{j=1}^{m} v_j\,.$$

One can deduce directly that any power n of this operator can be calculated recurrently as

$$\mathbf{Q}_\lambda^n(h)(z) = \max\Big(h(z),\, \mathbf{T}\big(\mathbf{Q}_\lambda^{n-1}(h)\big)(z) - \lambda\, g(z) \Big), \quad 1 \le n \le N\,.$$

Note that by the definition $\mathbf{Q}_\lambda^0(h) = h$. Now, we set the following stopping time

$$\mathbf{t}_\lambda^* = \min\left\{ 0 \le n \le N :\, \mathbf{Q}_\lambda^{N-n}(g)(Z_n) = g(Z_n) \right\}. \tag{9}$$

Finally, main contribution of this theory in following theorem.

Theorem 1. *For any $N \ge 1$ and for any $\lambda > 0$ the stopping time $\mathbf{t}_\lambda^*$ is the solution for the problem (5), i.e.*

$$\inf_{\tau \in \mathcal{M}} \Big(\lambda\, \widetilde{\mathbf{E}}(\tau - \nu)_+ + \widetilde{\mathbf{P}}(\tau < \nu) \Big) = \lambda\, \widetilde{\mathbf{E}}(\mathbf{t}_\lambda^* - \nu)_+ + \widetilde{\mathbf{P}}(\mathbf{t}_\lambda^* < \nu)\,.$$

Moreover, if there exists λ_α for which

$$\widetilde{\mathbf{P}}(\mathbf{t}_{\lambda_\alpha}^* \le \nu) = \alpha\,,$$

then the stopping time $\mathbf{t}_{\lambda_\alpha}^*$ is the solution for the problem (4), i.e.

$$\inf_{\tau \in \mathcal{M}_\alpha} \mathcal{R}(\tau) = \mathcal{R}(\mathbf{t}_{\lambda_\alpha}^*)\,.$$

Now through the Monte-Carlo simulations we compare the procedure (9) for the model (1) with $p = 1$ with the Bayesian rule from [9] defined as

$$T_B = \inf\{ 0 \le n \le N :\, \widetilde{\mathbf{P}}(\nu < n \,|\, X_1, \ldots, X_n) \ge 1 - \alpha \}\,, \quad 0 < \alpha < 1\,. \tag{10}$$

It should be noted that the posterior probability for the prior distribution can be represented as

$$\widetilde{\mathbf{P}}(\nu < n \,|\, X_1, \ldots, X_n) = \frac{(N+1)\,\overline{R}_n - 1}{(N+1)\,\overline{R}_n + N - n}\,.$$

Therefore, the procedure (10) can be rewritten as

$$T_B = \inf\left\{0 \le n \le N : \, (N+1)\,\overline{R}_n \ge \frac{N-n+1}{\alpha} - N + n\right\}. \tag{11}$$

In this case we use the mean delay risk defined as

$$\widehat{\mathcal{R}}(\tau) = \frac{1}{m\,(N+1)} \sum_{i=1}^{m} \sum_{k=0}^{N} \widehat{\mathcal{R}}_{k,i}(\tau) \quad \text{and} \quad \widehat{\mathcal{R}}_{k,i}(\tau) = \frac{1}{M} \sum_{l=1}^{M} (\tau^{l,i} - k)_+\,,$$

where $\tau^{l,i}$ is the lth sample of the stopping time τ with $\nu = k$ and $\theta = \theta_i$. Similarly we set the empiric probability of a false alarm as

$$\mathrm{PFA}(\tau) = \frac{1}{m\,(N+1)\,M} \sum_{i=1}^{m} \sum_{k=0}^{N} \sum_{l=1}^{M} \mathbf{1}_{\{\tau^{l,i} < k\}}\,,$$

where $\tau^{l,i}$ is the lth sample of the stopping time τ for the model (1) with $\nu = N$ and $\theta = \theta_i$. First we calculate this empiric probability for the procedure (9) and then we use this probability as α in (11). All calculations conducted for $\theta^* = 0$ and $\Theta = \{0.01, 0.03, 0.05\}$ with $M = 10^6$. The results are given in the Table 1.

Table 1. Results of Monte-Carlo simulations

N	λ	$\mathrm{PFA}(\mathbf{t}^*_\lambda)$	$\widehat{\mathcal{R}}(\mathbf{t}^*_\lambda)$	$\widehat{\mathcal{R}}(T_B)$	$\widehat{\mathcal{R}}(T_B)/\widehat{\mathcal{R}}(\mathbf{t}^*_\lambda)$
5	0.1000	$< 10^{-4}$	1.9987	1.9987	1.0000
5	0.1182	0.0001	1.9987	1.9990	1.0002
5	0.1455	0.1040	1.4940	2.0006	1.3391
5	0.3500	0.1667	1.1988	1.9994	1.6679
15	0.0100	$< 10^{-4}$	7.0019	7.0022	1.0001
15	0.0131	0.0066	6.8954	7.0039	1.0157
15	0.3500	0.0625	6.0590	6.9920	1.1540
15	0.4000	0.0625	6.0641	6.9974	1.1539
30	0.2000	0.0323	13.5206	14.4871	1.0715
30	0.3000	0.0323	13.5326	14.4994	1.0714
30	0.3500	0.0323	13.5350	14.5016	1.0714
30	0.4000	0.0323	13.5249	14.4915	1.0715

It should be noted that the Monte-Carlo simulations numerically confirm the theoretical results. As can be seen from Table 1, the optimal procedure (9) is uniformly better than the asymptotic Bayes procedure (10) for any $0 < \alpha < 1$. In the case when α is sufficiently small, i.e. $\alpha < 10^{-4}$, these procedures have the same average delay risks.

Acknowledgements. This research is partially supported by the RSF grant No 24-11-00191.

References

1. Arkoun, O., Brua, J-Yv., Pergamenshchikov, S.M.: Sequential Model Selection Method for Nonparametric Autoregression. Seq. Anal. **38**(4), 437–460 (2019)
2. Guépié, B.K., Lionel Fillatre, L., Nikiforov, I.: Sequential detection of transient changes. Seq. Anal. **31**, 528–547 (2012)
3. Lai, T.L.: Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. IEEE Trans. Inf. Theory **46**, 595–608 (2000)
4. Moustakides, G.V.: Multiple optimality properties of the Shewhart test. Seq. Anal. **33**, 318–344 (2014)
5. Pchelintsev, E.A., Pergamenchtchikov, S.M., Tenzin, R.O.: Truncated sequential change-point detection for Markov chains with applications in the epidemic statistical analysis. Seq. Anal. **43**(1–2), 57–78 (2024)
6. Pergamenchtchikov, S.M., Tartakovsky, A.G.: Asymptotically optimal point wise and minimax quickest change-point detection for dependent data. Stat. Inference Stoch. Process. **21**, 217–259 (2018)
7. Pergamenchtchikov, S.M., Tartakovsky, A.G.: Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis. J. Multivar. Anal. **174**, 104541 (2019)
8. Pergamenchtchikov, S.M., Tartakovsky, A.G., Spivak, V.: Minimax and pointwise sequential changepoint detection and identification for general stochastic models. J. Multivar. Anal. **190**, 104977 (2022)
9. Tartakovsky, A.G., Veeravalli, V.V.: General asymptotic Bayesian theory of quickest change detection. Theory Probab. Appl. **49**, 458–497 (2005)

Multi-way Contingency Tables with Uniform Margins

Elisa Perrone[1(✉)], Roberto Fontana[2], and Fabio Rapallo[3]

[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, Groene Loper 3, 5612 AE Eindhoven, The Netherlands
e.perrone@tue.nl

[2] Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
roberto.fontana@polito.it

[3] Dipartimento DIEC, Università di Genova, via Vivaldi 5, 16126 Genova, Italy
fabio.rapallo@unige.it

Abstract. We study the problem of transforming a multi-way contingency table into an equivalent table with uniform margins and same dependence structure. Such a problem relates to recent developments in copula modeling for discrete random vectors. Here, we focus on three-way binary tables and show that, even in such a simple case, the situation is quite different than for two-way tables. Many more constraints are needed to ensure a unique solution to the problem. Therefore, the uniqueness of the transformed table is subject to arbitrary choices of the practitioner. We illustrate the theory through some examples, and conclude with a discussion on the topic and future research directions.

Keywords: Categorical data analysis · odds ratios · multivariate Bernoulli · Iterative Proportional Fitting

1 Preliminaries

Tabular data in the form of contingency tables appear in many applications, such as health care, biology, and social science. Such data format have extensively been investigated in statistics with the primary goal of developing tools to extract information about the relations between the variables (see [7]). In [5], the author draws interesting connections between the analysis of bivariate contingency tables and *copulas*. We recall that the key concept behind copula theory is the separation between the marginal effect and the dependence structure. Such a separation allows for *ad hoc* modeling tools for dependence and is possible by transforming the original joint probability distribution into one with uniform margins on $[0, 1]$. This transformation can be obtained through the *Probability Integral Transform* (PIT). The transformed distribution is the copula associated with the original distribution and is uniquely defined if the margins are continuous [8]. In case of discrete random variables, the marginal

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 349–356, 2024.
https://doi.org/10.1007/978-3-031-65993-5_43

distributions cannot be transformed into uniform distributions through the PIT. In fact, the PIT only identifies the copula on a subdomain. As a consequence, there are infinite copula models that would fit such a subdomain. The question if we can adjust the copula theory to benefit from a similar idea of obtaining a margin-free model in the discrete context has been investigated in [5,6]. There and in the references therein, the authors explore the concept of transforming a given two-way contingency table into a new one with uniform margins. Such a transformation makes it easier to interpret the underlying association of the table which might be hidden when margins are heavily unbalanced, as discussed below.

For 2×2 contingency tables, the transformation entails converting the original table into a new one where all marginal probabilities equal 1/2, i.e., they are uniform on the subdomain. Table 1 reports a classical example from [9] where the data represents smallpox patients at Sheffield Hospital classified based on vaccination status [yes/no] and recovery [yes/no]. The odds ratio, which is a standard tool to measure association in contingency table analysis, is substantial as it equals 19.47. However, due to marginals that are far from uniform, the association may not be readily apparent in the original data. Nevertheless, the association becomes evident when the data is transformed to have marginal probabilities of 1/2 while maintaining the same odds ratio of 19.47. Figure 1 shows the effect of the transformation on the visualization of the data.

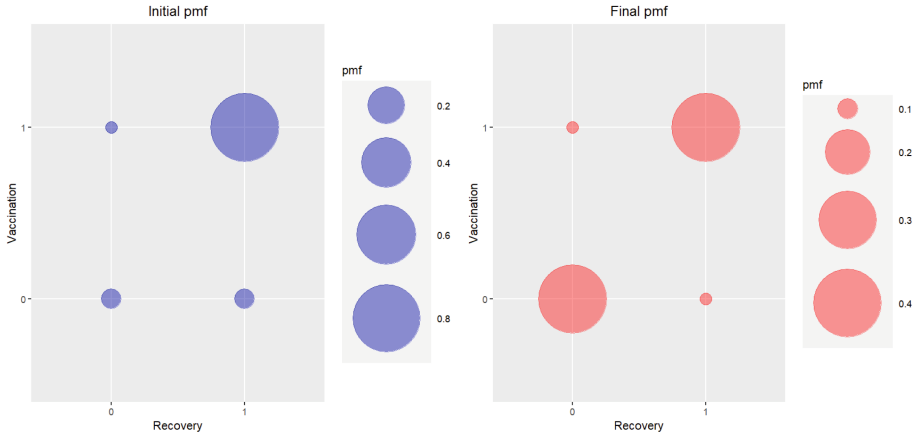
In this work, we investigate how to extend the idea of identifying an equivalent contingency table with uniform margins when we have dimension greater than two. In particular, we focus on the simplest case of $2 \times 2 \times 2$ contingency tables. In Sect. 2, we introduce the notation and show our main results. Some interesting examples are illustrated in Sect. 3, and we conclude with open questions for future research on the topic in Sect. 4.

Table 1. Sheffield smallpox epidemic reported in [9]

Vaccination (X_1)	Recovery (X_2)	$\tilde{n}$	$\tilde{p}$
No	No	274	0.06
No	Yes	278	0.06
Yes	No	200	0.04
Yes	Yes	3951	0.84

2 Multi-way Contingency Tables and Odds Ratios

We here outline the mathematical framework introduced by Geenens in [5] relevant for our work. We summarize the findings in two dimensions before moving the discussion to three dimensions.



(a) Original data (b) Transformed data

Fig. 1. Bubble plot of original data (left) and transformed data (right)

Two-way binary tables. Given a 2×2 contingency table $(\tilde{n}_{ij}, i, j = 0, 1)$ let $N = \tilde{n}_{00} + \tilde{n}_{11} + \tilde{n}_{01} + \tilde{n}_{10}$ be the grand total and $\omega_\star = \frac{\tilde{n}_{00}\tilde{n}_{11}}{\tilde{n}_{01}\tilde{n}_{10}}$ the odds ratio. For our goal, we consider the table written in terms of the relative frequencies $\tilde{p}_{ij} = \frac{\tilde{n}_{ij}}{N}$ instead of the counts $\tilde{n}_{ij}$. As explained in the motivating example of Sect. 1, the goal is to find a new table $(p_{ij} \geq 0, i, j = 0, 1)$ with marginals equal to $1/2$ that maintains the same odds ratio $\omega_\star$ of the original table. Such a table can be obtained by solving the system of equations in Eq. (1)

$$\begin{cases} \dfrac{p_{00}p_{11}}{p_{01}p_{10}} = \omega_\star \\ p_{00} + p_{01} + p_{10} + p_{11} = 1 \\ p_{00} + p_{01} - p_{10} - p_{11} = 0 \\ p_{00} - p_{01} + p_{10} - p_{11} = 0 \end{cases}, \tag{1}$$

whose unique solution only depends on the odds ratio and is as follows

$$\begin{cases} p_{00} = p_{11} = \dfrac{\sqrt{\omega_\star}}{2(1 + \sqrt{\omega_\star})} \\ p_{01} = p_{10} = \dfrac{1}{2(1 + \sqrt{\omega_\star})} \end{cases}. \tag{2}$$

Using the data in Table 1, we obtain $p_{00} = p_{11} \approx 0.41$ and $p_{01} = p_{10} \approx 0.09$.

Three-way binary tables. We now consider three classification variables X_1, X_2, X_3, with relative frequencies denoted by $p_{i,j,k}$, for $i, j, k \in \{0, 1\}$. A possible generalization of the approach presented for two-way tables to three-way tables goes through the definition of the *3-dimensional odds ratio* ω_3 given as

$$\omega_3 = \frac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}p_{111}}. \tag{3}$$

Additionally, one can also define the so-called *conditional odds ratios*:

$$\omega_2(X_i, X_j|X_k = a),\ i \neq j \neq k \tag{4}$$

where $\omega_2(X_i, X_j|X_k = a)$ is the 2-dimensional odds ratio computed for X_i and X_j referring to the sub-table defined by $X_k = a$, $a \in \{0, 1\}$. Since k is uniquely determined by i and j, we streamline the notation by writing $\omega_{ij|a}$ instead of $\omega_2(X_i, X_j|X_k = a)$. Interestingly, the 3-dimensional odds ratio and the conditional odds ratios are linked by the following chain of equalities (see [7]):

$$\omega_3 = \frac{\omega_{12|0}}{\omega_{12|1}} = \frac{\omega_{23|0}}{\omega_{23|1}} = \frac{\omega_{13|0}}{\omega_{13|1}}. \tag{5}$$

We are now ready to propose a possible generalization of the problem presented for two-way binary tables. In particular, we consider a $2 \times 2 \times 2$ contingency table $(\tilde{n}_{ijk}, i, j, k = 0, 1)$, with grand total N, relative frequencies $\tilde{p}_{ijk} = \tilde{n}_{ijk}/N, i, j, k \in \{0, 1\}$, 3-dimensional odds ratio $\omega_\star$ (defined as in Eq. (3)), and conditional odds ratios $\omega^\star_{ij|a}$ (defined as in Eq. (4)). The goal is to find the relative frequencies of a new table $(p_{ijk} \geq 0, i, j, k \in \{0, 1\})$ with uniform marginals equal to $1/2$ that maintains the same odds ratio $\omega_\star$ as the original table. Basically, we transform the problem of Eq. (1) into the system of equations reported below:

$$\begin{cases} \dfrac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}p_{111}} = \omega_\star \\ p_{000} + p_{001} + p_{010} + p_{011} + p_{100} + p_{101} + p_{110} + p_{111} = 1 \\ p_{000} + p_{001} + p_{010} + p_{011} - p_{100} - p_{101} - p_{110} - p_{111} = 0 \\ p_{000} + p_{001} - p_{010} - p_{011} + p_{100} + p_{101} - p_{110} - p_{111} = 0 \\ p_{000} - p_{001} + p_{010} - p_{011} + p_{100} - p_{101} + p_{110} - p_{111} = 0 \end{cases}. \tag{6}$$

Differently from the case of two-way tables, the solution of the system in Eq. (6) is not unique. One possible solution is given by the generalization of the 2-dimensional solution Eq. (2) as follows

$$\mathbf{p}_\star = \begin{cases} p_{000} = p_{011} = p_{101} = p_{110} = \dfrac{\sqrt[4]{\omega_\star}}{4(1 + \sqrt[4]{\omega_\star})} \\ p_{111} = p_{100} = p_{010} = p_{001} = \dfrac{1}{4(1 + \sqrt[4]{\omega_\star})} \end{cases}. \tag{7}$$

However, there are in general many more solutions suggesting that this straightforward extension is not enough to guarantee a prescribed dependence structure in the new table. A way to ensure a unique solution is to add more constraints to the problem. In particular, a natural choice would be to require that the transformed table p_{ijk} preserves also the conditional odds ratios of the original table. A new system of equations that includes the additional constraints is given in Eq. (8). We notice that the relationships highlighted in Eq. (5) allows us to only use the conditional odds ratios $\omega^\star_{23|0}, \omega^\star_{13|0}$, and $\omega^\star_{12|0}$, since the others can

be derived by combining them with the 3-dimensional odds ratio. The problem presented in Eq. (8) has a unique solution that can be computed using symbolic mathematical software like Mathematica

$$\begin{cases} \dfrac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}p_{111}} = \omega_\star \\ \dfrac{p_{000}p_{011}}{p_{001}p_{010}} = \omega^\star_{23|0} \\ \dfrac{p_{000}p_{101}}{p_{001}p_{100}} = \omega^\star_{13|0} \\ \dfrac{p_{000}p_{110}}{p_{010}p_{100}} = \omega^\star_{12|0} \\ p_{000}+p_{001}+p_{010}+p_{011}+p_{100}+p_{101}+p_{110}+p_{111}=1 \\ p_{000}+p_{001}+p_{010}+p_{011}-p_{100}-p_{101}-p_{110}-p_{111}=0 \\ p_{000}+p_{001}-p_{010}-p_{011}+p_{100}+p_{101}-p_{110}-p_{111}=0 \\ p_{000}-p_{001}+p_{010}-p_{011}+p_{100}-p_{101}+p_{110}-p_{111}=0 \end{cases} . \tag{8}$$

3 Examples

In this section, we present two examples to illustrate the results reported in Sect. 2. First, we analyze the artificial data of Table 2, taken from [1]. The cross-classification involves three binary variables, which are the response to a treatment (success or failure, X_1), the treatment (drug A or drug B, X_2), and the clinic (1 or 2, X_3). The frequencies in Table 2 are expected counts from a model with conditional odds ratio $\omega^\star_{12|0}$ (treatment and response given the clinic) equal to 1. The other conditional odds-ratios are given by $\omega^\star_{23|0} = 6$, $\omega^\star_{13|0} = 0.167$, and the 3-dimensional odds-ratio $\omega_\star = 1$. Through Mathematica, it was possible to find the solution p to the problem of Eq. (8). The solution is reported in Table 2. Again using Mathematica, we were also able to find the closed form solution to the more general problem of Eq. (6) with the only constraint (besides uniform margins) that the 3-dimensional odds ratio equals 1. The solution has three free variables and is reported in the following Eq. (9):

$$\begin{cases} 0 < p_{000} < 1/2 \\ 0 < p_{001} < 1/2(1-2p_{000}) \\ 0 < p_{010} < 1/2(1-2p_{000}-2p_{001}) \\ p_{011} = 1/2(1-2p_{000}-2p_{001}-2p_{010}) \\ p_{100} = p_{011} \\ p_{101} = 1/2(2p_{010}+2p_{011}-2p_{100}) \\ p_{110} = 1/2(2p_{001}+2p_{011}-2p_{100}) \\ p_{111} = p_{000}+p_{001}+p_{010}+p_{011}-p_{100}-p_{101}-p_{110} \end{cases} . \tag{9}$$

We now consider another example that is based on real data reported in Table 3. Such dataset has been analyzed in [3] to illustrate interactions in three-

Table 2. Response versus treatment and clinic, from [1]. Counts and solution to problem Eq. (8) (in brackets)

		Response (X_1)	
Clinic (X_3)	Treatment (X_2)	Success	Failure
1	A	18	12
		(0.252)	(0.103)
	B	12	8
		(0.103)	(0.042)
2	A	2	8
		(0.042)	(0.103)
	B	8	32
		(0.103)	(0.252)

way tables. In this example, thirty patients suffering from lymphocytic lymphoma who responded to a course of combination chemotherapy are classified by sex (X_2) and cell type (X_3). With a log-linear analysis, the author concludes that the cell type is related to both sex and outcome of the therapy (X_1). By using Mathematica, we could find the solution of the constrained problem with same odds ratio structure of the original data but uniform margins. The new table entries are showed in Table 3.

Table 3. Response of Lymphoma patients to Combination Chemotherapy: Distribution by Sex and Cell type, from [3]. Counts and solution to problem Eq. (8) (in brackets)

		Outcome (X_1)	
Cell type (X_3)	Sex (X_2)	No Response	Response
Nodular	Male	1	4
		(0.024)	(0.133)
	Female	2	6
		(0.065)	(0.278)
Diffuse	Male	12	1
		(0.305)	(0.040)
	Female	3	1
		(0.105)	(0.050)

In these examples the difference between the original table and the transformed one is less evident because the three marginal distributions are more balanced than those in Table 1. We can conclude that the more unbalanced the margins, the more advantageous the transformation, as it provides more insights on the dependence structure. We also notice that the solution derived

with our method coincides with the one obtained by applying the standard Iterative Proportional Fitting Procedure (IPFP) [2,7]. The connection between the two approaches will be subject of further studies.

4 Conclusions and Discussion

In this work, we present possible extensions of [5] to multi-way contingency tables. The primary goal of our investigation is to derive a unique solution to the problem of searching for a new table with preserved dependence structure and uniform margins. As discussed in Sect. 2 and Sect. 3, there are infinite tables with uniform margins and a fixed 3-dimensional odds ratio. However, we show that the uniqueness is guaranteed when also imposing constraints on the conditional odds-ratios. We notice that this is not the only possible choice to ensure uniqueness. For example, we could add constraints on the uniformity of the 2-dimensional sections of the multi-way table. Such a choice results in searching for the probabilities $p_{ijk} \geq 0,\ i,j,k = 0,1$ that satisfy the system of equations reported in the following Eq (10):

$$\begin{cases} \dfrac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}p_{111}} = \omega_\star \\ p_{000}+p_{001}+p_{010}+p_{011}+p_{100}+p_{101}+p_{110}+p_{111} = 1 \\ p_{000}+p_{001}+p_{010}+p_{011}-p_{100}-p_{101}-p_{110}-p_{111} = 0 \\ p_{000}+p_{001}-p_{010}-p_{011}+p_{100}+p_{101}-p_{110}-p_{111} = 0 \\ p_{000}-p_{001}+p_{010}-p_{011}+p_{100}-p_{101}+p_{110}-p_{111} = 0 \\ p_{000}+p_{001}-p_{010}-p_{011} = 0 \\ p_{010}+p_{011}-p_{100}-p_{101} = 0 \\ p_{100}+p_{101}-p_{110}-p_{111} = 0 \\ p_{000}-p_{001}+p_{010}-p_{011} = 0 \\ p_{001}+p_{011}-p_{100}-p_{110} = 0 \\ p_{100}-p_{101}+p_{110}-p_{111} = 0 \\ p_{000}-p_{001}+p_{100}-p_{101} = 0 \\ p_{001}-p_{010}+p_{101}-p_{110} = 0 \\ p_{010}-p_{011}+p_{110}-p_{111} = 0 \end{cases} . \tag{10}$$

Interestingly, the unique solution to this problem is the generalization of Eq (2) given by the following entries of the three-way table

$$\begin{cases} p_{000} = p_{011} = p_{101} = p_{110} = \dfrac{\sqrt[4]{\omega_\star}}{4(1+\sqrt[4]{\omega_\star})} \\ p_{111} = p_{100} = p_{010} = p_{001} = \dfrac{1}{4(1+\sqrt[4]{\omega_\star})} \end{cases} .$$

Although the solution is unique, this approach results in a constrained dependence of (X_1, X_2), (X_1, X_3), and (X_2, X_3) which is then modified from the original table structure. Therefore, it is not in line with the idea of maintaining

the dependence structure of the original table. Unique solutions could also be obtained under other types of stochastic constraints, such as imposing exchangeability. We will investigate this aspect in the future. Additionally, we plan to extend the theory to arbitrary three-way tables. We believe that this extension is always possible if there are no zeros in the table and the (conditional) odds ratios can be computed. If zero entries are present, there might be cases where there is no table with uniform margins that also maintains the odds-ratio structure. Extending Thm. 6.1 of [5] to multi-way tables is needed to ensure the existence of such a table. In future work, we plan to analyze this aspect and build on previous results on the impact of structural zeros for two-way tables presented in [4]. Finally, extending the approach introduced here to arbitrary multi-way contingency tables in comparison with existing methods, i.e., IPFP, is also an interesting direction for further research. Such a case is non trivial: as the dimension grows, the number of conditional odds ratios and the degree of the involved equations grow exponentially. A possible way to ensure the feasibility of the problem is to pair odds-ratio constraints with stochastic constraints that naturally reduce the space of solutions. We will explore this idea in a follow-up paper.

Acknowledgments. We thank the anonymous reviewer for their comments on a previous version of the manuscript. Additionally, Roberto Fontana and Fabio Rapallo are members of the GNAMPA-INdAM group.

References

1. Agresti, A.: Categorical Data Analysis. 3rd edn. Wiley (2012)
2. Barthélemy, J., Suesse, T.: mipfp: an R package for multidimensional array fitting and simulating multivariate Bernoulli distributions. J. Stat. Softw. **86**(2), 1–20 (2018)
3. Bishop, Y.M., Fienberg, S.E., Holland, P.W.: Discrete Multivariate Analysis. Springer (2007)
4. Fontana, R., Perrone, E., Rapallo, F.: Contingency tables with structural zeros and discrete copulas. In: Book of the Short Papers - SIS 2023, pp. 713–718. Pearson (2023)
5. Geenens, G.: Copula modeling for discrete random vectors. Depend. Model. **8**(1), 417–440 (2020)
6. Kojadinovic, I., Martini, T.: Copula-like inference for discrete bivariate distributions with rectangular supports. arXiv:2307.04225 (2024)
7. Rudas, T.: Lectures on Categorical Data Analysis. Springer (2018)
8. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)
9. Yule, G.U.: On the methods of measuring association between two attributes. J. R. Stat. Soc. **75**(6), 579–652 (1912)

Low-Rank Approximation of Data Matrices Using Robust Sparse Principal Component Analysis

Pia Pfeiffer(✉) and Peter Filzmoser

TU Wien, 1040 Vienna, Austria
{pia.pfeiffer,Peter.Filzmoser}@tuwien.ac.at

Abstract. The estimation of principal components can be influenced by outlying observations, so-called row-wise outliers. For high-dimensional data, it becomes more and more likely that an observation contains outlying cells, which would lead to many row-wise outliers and to a breakdown of traditional robust methods. In this case, it is preferable to achieve protection against cell-wise outliers. We present various approaches for principal component analysis that lead to row-wise and cell-wise robustness. Moreover, we focus on sparse methods that enforce zeros in the loadings matrix and thus simplify the interpretation.

Keywords: cell-wise robustness · sparsity · multivariate statistics

1 Introduction

Principal Component Analysis (PCA) might not be considered the most high-level machine learning method, as it dates back to a period, when such "machines" did not even exist [13,22]. Nevertheless, PCA is still widely used as a first step in multivariate data analysis, simply to get an overview of the inherent data structure. As the first few principal components (PCs) refer to the main variability of the data, they should contain valuable information on the relationships among the variables, and about possible groups among the observations. Because of these properties, the first few PCs can also be used to reconstruct the data matrix. In fact, the well-known Eckhart-Young theorem says that the first k PCs give the best rank-k approximation of a data matrix with respect to the Frobenius norm [9]. Low-rank approximations, especially in the context of SVD (Singular Value Decomposition), are also widely discussed in the more recent literature, in particular in the machine learning community (e.g. [18]). With increasing size of data sets, this leads to huge advantages in data storage, and it is a way to filter out noise and thus focus on the main data information.

The Frobenius norm as an L_2 norm is sensitive to outliers in the data matrix. As a consequence, the resulting rank-k approximation could in the worst case just reveal information about the outliers, but not about the main data structure.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 357–362, 2024.
https://doi.org/10.1007/978-3-031-65993-5_44

Therefore, approaches exist where the Frobenius norm has been replaced by other norms or distance measures to give less weight to outlying observations [18,20]. Equivalently, this leads to robust PCs, being less affected by outliers.

Historically, robustness refers to outlying (multivariate) observations [21]. This goes back to the so-called epsilon-contamination model, where it is assumed that the majority of the observations is generated from a "clean" distribution, while a minority originates from a contaminated (unspecified) distribution [21]. With increasing dimensionality of the data sets, this concept has been extended to cell-wise contamination, where not the complete observation but single cells of a data matrix are considered as outliers [1]. The usefulness in this context is clear: If outlying cells are spread across the whole high-dimensional data matrix, it could happen that most of the observations contain contaminated cells. In the traditional concept of robustness, these observations would be treated as outliers, which then form the data majority, and would cause the breakdown of any traditional robust estimator. For this reason, so-called cell-wise robust estimators have been introduced in the literature, starting with outlier detection [25], and extending to several other methods such as regression (e.g. [11]) and PCA [16].

Next to robustness, a further desirable property for PCA, especially in the high-dimensional setting, is sparsity of the loadings matrix, referring to zero entries in this matrix. Zeros exclude contributions of the corresponding variables to the PCs and thus simplify the interpretation of the PCs. Robust sparse PCA has been introduced in [8], and several alternative proposals followed.

The aim of this contribution is to provide an overview of approaches to robust PCA, sparse robust PCA, cell-wise robustness for PCA, and the combination of cell-wise robust and sparse PCA.

2 Robust PCA

There are various different ways how to robustify PCA. In the following, we focus on two approaches, which seem to be the main streams in robust PCA.

2.1 Robust PCA Based on Robust Covariance Estimation

Let $\boldsymbol{X}$ be a data matrix with n observations forming the rows, and p variables forming the columns. We assume that the columns are mean-centered by any (robust) estimator of the mean. The $n \times p$ matrix $\boldsymbol{Z}$ of PC scores can be obtained by the linear combinations $\boldsymbol{Z} = \boldsymbol{XV}$, where the $p \times p$ loadings matrix $\boldsymbol{V}$ is derived from the eigen decomposition of the estimated covariance matrix $\boldsymbol{C}$ of $\boldsymbol{X}$, thus

$$\boldsymbol{C} = \boldsymbol{VAV}', \tag{1}$$

where $\boldsymbol{V}$ contains the normed eigen vectors in its columns, and $\boldsymbol{A} = \mathrm{Diag}(a_1, \ldots, a_p)$ is a diagonal matrix with the corresponding eigenvalues sorted in ascending order. Classical, non-robust PCs would be obtained by taking the sample covariance matrix as the matrix $\boldsymbol{C}$ in Eq. (1). If instead a robust covariance matrix is used, the resulting PCs are robust. The robustness has been characterized

by the influence functions and asymptotic efficiencies for the eigenvectors and eigenvalues [3] of several covariance estimators, such as the M-estimator [19], the S-estimator [6], the MCD-estimator [24], the spatial sign covariance estimator [5], or the MM estimator [26].

2.2 Robust PCA Based on Projection-Pursuit

While the plug-in approach mentioned in the last section is simple and leads to desirable robustness properties, these estimators cannot be computed when the number of variables p is bigger than the number of observations n. To handle this situation, [4] introduced the projection-pursuit (PP) method for robust PCA. In order to find the first PC, one seeks a normed vector $\boldsymbol{b}_1$ such that the variance of $\boldsymbol{X}\boldsymbol{b}_1$ is maximized. Higher-order PCs are defined similarly, by identifying normed vectors $\boldsymbol{b}_j$ that maximize the variance of $\boldsymbol{X}\boldsymbol{b}_j$ under the constraints $\boldsymbol{b}_j'\boldsymbol{b}_k = 0$ for $j > k$. If the variance estimator is the sample variance, the resulting vectors are the eigenvectors of the sample covariance matrix, see Eq. (1). However, other robust estimators of variance (or scale), such as the MAD or the Qn, lead to robust PCs [4]. Thus, this approach avoids the estimation of a covariance matrix but simplifies the problem to univariate scale estimation. Even more, since the PCs are estimated sequentially, one could stop at a desired number of PCs, which is a clear advantage in the high-dimensional setting. Numerical algorithms have been proposed and compared in [7].

A method combining both the projection-pursuit and the plug-in approach is ROBPCA [14]. The essential steps are first projecting the data to a low-dimensional subspace and then applying a robust estimator of the covariance.

3 Robust and Sparse PCA

Especially for high-dimensional data, the interpretation of PCs often becomes challenging, since potentially many variables contribute to the individual PCs. This is simplified by sparse PCA, where some entries of the loadings matrix are set to zero.

A first proposal in this direction was the SCoTLASS (Simplified Component Technique - LASSO) method, which adds an L_1 penalty as a constraint to the optimization problem [17]. Thus, one looks for a normed vector $\boldsymbol{b}_1$ such that $\boldsymbol{b}_1'\boldsymbol{C}\boldsymbol{b}_1$ is maximized under the constraint $\|\boldsymbol{b}_1\|_1 \leq t$, for a constant t. Here, $\boldsymbol{C}$ is the estimated covariance matrix. The solution $\hat{\boldsymbol{b}}_1$ is the first (sparse) loadings vector to the scores $\boldsymbol{X}\hat{\boldsymbol{b}}_1$. Higher-order PCs are obtained similarly, by imposing orthogonality constraints to the previous direction vectors. This non-convex optimization problem can be solved by a gradient-ascent algorithm [17].

In principle, one could use a plug-in approach, similar to Sect. 2.1, to plug in a robustly estimated covariance matrix into the optimization problem and obtain robust PCs. However, we would face the same limitations as previously for high-dimensional data.

It is thus natural to consider the approach of Sect. 2.2, and add an L_1 penalty. In more detail, the aim is to find the normed vectors $\boldsymbol{b}_j$ that maximize the

(robust) variance of $\boldsymbol{X}\boldsymbol{b}_j$, under the constraint $\|\boldsymbol{b}_j\|_1 \leq t$, and orthogonality to previously estimated direction vectors. This was proposed in [8], together with an iterative algorithm to compute the solution.

Several other robust and sparse PCA methods have been proposed in the literature. For example, [15] proposed the method ROSPCA, an extension to their earlier work ROBPCA [14], which integrates an L_1 penalty. [12] proposed a robust extension of sparse PCA via regression [29] that is based on robust plug-in estimators for the covariance.

4 Cell-Wise Robust PCA

All previously mentioned robust PCA methods can deal with row-wise outliers. Especially in the high-dimensional setting, however, robustness against cell-wise outliers is preferable [23]. For cell-wise outliers, the information in the majority of the variables can be valuable, so down-weighting the whole row would lead to a substantial loss of information.

A cell-wise robust PCA method is macroPCA [16], and it combines the detecting-deviating-cells (DDC) algorithm [25] with a version of ROBPCA [14].

In the following, we focus on low-rank approximation based on PCA. Suppose that $\boldsymbol{Z}_k$ contains the first k PC scores in its columns and $\boldsymbol{V}_k$ the corresponding direction vectors (loadings) of the first k PCs. Then $\hat{\boldsymbol{X}} = \boldsymbol{Z}_k\boldsymbol{V}_k'$ is the rank-k approximation of the data matrix $\boldsymbol{X}$, and $\boldsymbol{R} = \boldsymbol{X} - \hat{\boldsymbol{X}}$ is the residual matrix of this approximation, with residuals r_{ij}. Minimizing the (squared) Frobenius norm $\|\boldsymbol{R}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p r_{ij}^2$ yields the classical PCs as solutions, being sensitive to outliers. An alternative formulation of the problem is

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \rho\left(\frac{r_{ij}}{\hat{\sigma}_j}\right) = \min, \tag{2}$$

where $\hat{\sigma}_j$ is a scale estimate of the j-th variable [27,28]. In order to achieve good robustness properties, the function ρ needs to be bounded. [20] proposed a procedure based on an alternating regression algorithm to obtain cell-wise robust PCs.

5 Cell-Wise Robust and Sparse PCA

The idea used in Eq. (2) can be modified by adding an L_1 penalty to the loading vectors, as done in Sect. 3. Naturally, this leads to a PCA version that is cell-wise robust and sparse. One could adjust the sparsity for each loading vector by using separate tuning parameters, or, alternatively, one can regulate the overall sparsity in the loadings with one common tuning parameter.

The resulting optimization problem is treated in the framework of optimization on manifolds [10]. In particular, stochastic gradient descent algorithms can be extended when the objective function is defined on a Riemannian manifold, and convergence properties can be derived [2]. This also allows for the scalability

of the algorithm to datasets containing a very large number of observations. The implementation of the procedure is available at https://github.com/piapfeiffer/RobSparseMVA/.

In the talk, we will present the algorithm to solve this complex optimization problem in more detail, and demonstrate the usefulness of this method by real data applications and simulation results, which reveal the superiority of this approach to existing PCA methods.

6 Summary and Conclusions

PCA belongs to the most important methods in multivariate data analysis. Since the goal is to uncover the main structure inherent in the multivariate data, it is desirable that outliers do not mislead the method by artificially inflating the variance. For so-called flat data, having more variables than observations, it is even more challenging to focus on the main data structure if outlying cells are spread across the whole data matrix. Thus, cell-wise robust methods are advantageous in such a setting, but also sparsity brings benefits with respect to the interpretation of the resulting components. In the paper we presented various approaches on the way to a cell-wise robust and sparse PCA method. The results can be further explored in several different ways: It could be interesting to perform outlier diagnostics or to proceed with the resulting PCs for regression or classification.

Acknowledgments. This work was funded by the Austrian COMET-Program (project InTribology2, No. 906860) via the Austrian Research Promotion Agency (FFG) and the federal states of Niederösterreich and Vorarlberg and was carried out at the Austrian Excellence Centre of Tribology (AC2T research GmbH) and the TU Wien.

References

1. Alqallaf, F., Van Aelst, S., Yohai, V.J., Zamar, R.H.: Propagation of outliers in multivariate data. Ann. Stat. **37**(1), 311–331 (2009)
2. Bonnabel, S.: Stochastic gradient descent on Riemannian manifolds. IEEE Trans. Autom. Control **58**(9), 2217–2229 (2013)
3. Croux, C., Haesbroeck, G.: Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. Biometrika **87**(3), 603–618 (2000)
4. Croux, C., Ruiz-Gazen, A.: High breakdown estimators for principal components: the projection-pursuit approach revisited. J. Multivar. Anal. **95**(1), 206–226 (2005)
5. Croux, C., Ollila, E., Oja, H.: Sign and rank covariance matrices: statistical properties and application to principal components analysis. In: Dodge, Y. (ed.) Statistical Data Analysis Based on the L1-Norm and Related Methods, pp. 257–269. Birkhäuser Basel, Basel (2002)
6. Croux, C., Van Aelst, S., Dehon, C.: Bounded influence regression using high breakdown scatter matrices. Ann. Inst. Stat. Math. **55**, 265–285 (2003)

7. Croux, C., Filzmoser, P., Oliveira, M.R.: Algorithms for projection-pursuit robust principal component analysis. Chemometr. Intell. Lab. Syst. **87**(2), 218–225 (2007)
8. Croux, C., Filzmoser, P., Fritz, H.: Robust sparse principal component analysis. Technometrics **55**(2), 202–214 (2013)
9. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychometrika **1**(3), 211–218 (1936)
10. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998)
11. Filzmoser, P., Höppner, S., Ortner, I., Serneels, S., Verdonck, T.: Cellwise robust M regression. Comput. Stat. Data Anal. **147**, 106944 (2020)
12. Greco, L., Farcomeni, A.: A plug-in approach to sparse and robust principal component analysis. TEST **25**, 449–481 (2016)
13. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. **24**(6), 417 (1933)
14. Hubert, M., Rousseeuw, P.J., Vanden Branden, K.: ROBPCA: a new approach to robust principal component analysis. Technometrics **47**(1), 64–79 (2005)
15. Hubert, M., Reynkens, T., Schmitt, E., Verdonck, T.: Sparse PCA for high-dimensional data with outliers. Technometrics **58**, 424–434 (2016)
16. Hubert, M., Rousseeuw, P.J., Van den Bossche, W.: MacroPCA: an all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers. Technometrics **61**(4), 459–473 (2019)
17. Jolliffe, I.T., Trendafilov, N.T., Uddin, M.: A modified principal component technique based on the LASSO. J. Comput. Graph. Stat. **12**(3), 531–547 (2003)
18. Li, Y., Woodruff, D.: Input-sparsity low rank approximation in Schatten norm. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 6001–6009 (2020)
19. Maronna, R.A.: Robust M-estimators of multivariate location and scatter. Ann. Stat. **4**(1), 51–67 (1976)
20. Maronna, R.A., Yohai, V.J.: Robust low-rank approximation of data matrices with elementwise contamination. Technometrics **50**(3), 295–304 (2008)
21. Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M.: Robust Statistics: Theory and Methods (with R). Wiley (2019)
22. Pearson, K.: On lines and planes of closest fit to systems of points in space. Lond. Edinburgh Dublin Philos. Mag. J. Sci. 2(11), 559–572 (1901)
23. Raymaekers, J., Rousseeuw, P.J.: Challenges of cellwise outliers. arXiv preprint arXiv:2302.02156 (2023)
24. Rousseeuw, P.J.: Multivariate estimation with high breakdown point. Math. Stat. Appl. **8**, 283–297 (1985)
25. Rousseeuw, P.J., Bossche, W.V.D.: Detecting deviating data cells. Technometrics **60**(2), 135–145 (2018)
26. Salibián-Barrera, M., Van Aelst, S., Willems, G.: Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. J. Am. Stat. Assoc. **101**(475), 1198–1211 (2006)
27. De la Torre, F., Black, M.J.: Robust principal component analysis for computer vision. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 1, pp. 362–369. IEEE (2001)
28. Verboon, P., Heiser, W.J.: Resistant lower rank approximation of matrices by iterative majorization. Comput. Stat. Data Anal. **18**(4), 457–467 (1994)
29. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. 15(2), 265–286 (2006)

Multi-agent Reinforcement Learning and Its Application to Wireless Network Communication

Sabrina Pochaba[1,2(✉)], Peter Dorfinger[1], Matthias Herlich[1], Roland Kwitt[2], and Simon Hirlaender[2]

[1] Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria
[2] Paris-Lodron University, Salzburg, Austria
sabrina.pochaba@salzburgresearch.at

Abstract. Real-world problems often involve many players interacting with each other while pursuing their own goals. Although the topic of Machine Learning (ML) is making great progress in picturing good learning strategies, there is a lack of consideration of many player strategies. Thus, the topic of Multi-Agent Reinforcement Learning (MARL) attempts to address this need. Here, the learning strategies of Reinforcement Learning (RL), where one agent interacts with an environment to solve a task, are extended to multiple players trying to solve their own tasks. To do so, algorithms from RL are used and applied to multi-agent settings, coupled with game-theoretic aspects to predict the multi-player behaviour. After explaining the main concepts of MARL with their challenges and advantages, we apply MARL to a setting of wireless communication. Here, the multi-agent setting can take into account the different number of communicating devices that communicate in real-world communication scenarios.

Keywords: Reinforcement learning · Multi-agent reinforcement learning · Game theory · Wireless network communication · Device-to-device communication

1 Introduction

When real-world problems should be solved with ML, there can occur the difficulty of many player problems, where different players decide what to do by achieving their own goals without consideration of other players. To get access to those difficult problem definitions, a mixture of game theory and ML can be a good approach to face this challenge. Since the interaction of many different players in such problems influence each other without really noticing, it is often not sufficient to use algorithms that train only one player in his specific scenario. With the approach of MARL, we get algorithms that can face multi-player scenarios. In this approach, many players learn to solve their task by

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 363–370, 2024.
https://doi.org/10.1007/978-3-031-65993-5_45

interacting with each other. This interaction of many players comes with some more challenges, that need to be addressed, when using MARL.

In wireless communication, where many devices communicate at the same time, we have some many-player problems as well. Thus, it is suitable to use MARL to assign the resources of these devices properly. We also use game theoretical approaches to get a better understanding of the underlying scenario.

In this paper we discuss the main subject of MARL, as well as the underlying challenges and apply MARL to a specific problem in wireless network communication. Thus, the paper is structured as follows:

Section 2 explains the basic concept of RL, which is then exctended to MARL in Chap. 3. The following Sect. 4 explores game theory and how it can be applied to MARL. Afterwards, Sect. 5 shows the application of MARL to a problem of resource allocation in wireless network communication. Then, the paper is completed with a chapter combining future work and the conclusion.

2 Reinforcement Learning

RL is a subject of Machine Learning, where an agent learns to solve a task by trial and error search. To do so, the agent interacts with an environment that represents the problem. This environment gives a state s_t to the agent, that shows the agent what he knows from the actual state of the environment. Based on this state, the agent takes an action a_t, trying to solve his task. This action then changes the environment and the environment returns its new state s_{t+1} to the agent. In order that the agent can learn, the environment additionally gives a reward r_t to the agent. This reward tells the agent, how good his behaviour was. By going through this loop between states and rewards of the environment and actions of the agent, the agent learns its task by maximizing the cumulative reward [5].

During training, the agent improves his policy π. This policy describes a function from the state space S to the action space A and aims to maximize the cumulative reward, which is called the return G_t given by

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}.$$

The agent maximizes this return by calculating the expectation value of it in regard to the state and the action, which is called Q-Function $q_\pi(s, a)$. The Q-Function is defined as

$$q_\pi(s, a) = \mathbf{E}_\pi \left[G_t | s_t = s, a_t = a \right] \ , \ s \in S, a \in A,$$

which characterizes the expectation of the return starting from a fixed state $s \in S$ and taking action $a \in A$ [5].

Using the expected return in the (State-)Value-Function or the Q-Function, there occur many different RL methods converging to an optimal policy π^*. However, those RL methods need the prerequisite that the task is embedded

in a Markov Decision Process. That means, that all states have to be Markov, which says that all states have to contain all information of the past that is needed for solving the task.

3 Multi-agent Reinforcement Learning

MARL is an extension of RL (see Fig. 1). It describes a scenario, where not just one, but many agents learn simultaneously to solve a task. Thus, the agents influence each other because they execute their actions in the same environment. In a MARL setting with n agents, the action $a_t = (a_t^1, a_t^2, ..., a_t^n)$ is a joint action, collecting the actions of all agents $j = 1, ..., n$. The same holds for the state $s_t = (s_t^1, s_t^2, ..., s_t^n)$, that is a joint state of the agent's states. Thus, we have a joint state space $S = (S^1, S^2, ..., S^n)$ and joint action space $A = (A^1, A^2, ..., A^n)$. The reward can be shown as a joint reward $r_t = (r_t^1, r_t^2, ..., r_t^n)$ as well [6].

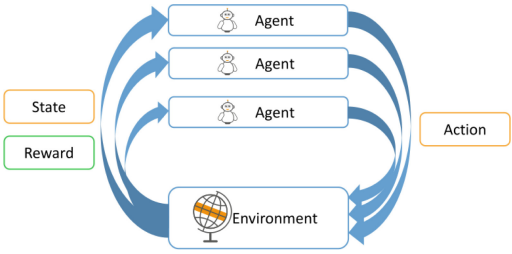


Fig. 1. In MARL the environment gives a state and a reward to all agents. The agents take their actions based on their state and thus, changing the environment.

With Multi-Agent settings, it is possible to model many natural problems. The easiest MARL task is a group of cooperative agents. This can be achieved by selecting a shared learning reward, that should be maximized. Another kind of tasks occur if the agents act in an adversarial manner. In that case, every agent follows his own goal. The ignorant behaviour of adversarial agents makes it more difficult to train them. However, there is also a mixed variant of the above-mentioned behaviour with different groups of agents, where agents act cooperative within one group, but in an adversarial manner against the agents of other groups. An example of those mixed settings can be found in hide-and-seek games [2].

3.1 Challenges of MARL

The subject of MARL is associated with some challenges. One major difficulty is the scalability of the state and the action space. Due to the many agents, the size of these spaces increase exponentially. These large state and action spaces can lead to computational problems [7].

Another challenge of MARL, is the information structure of the scenario. Depending on whether the setting is cooperative or adversarial, there can occur different information structures. In some settings, there exist a central controller, that communicates with the agents and exchanges information with them. However, a central controller often does not exist in real world problems. For applicable real world problems it is more realistic to concentrate either on a decentralized setting with networked agents, where the agents can exchange information, or on a fully decentralized setting. The decentralized setting with networked agents is easier to solve, due to the interaction of the agents. Since communication of the agents is not always the case, it is most likely that a real world Multi-Agent system builds a fully decentralized setting, where the agents cannot communicate with each other and are often not even aware of the presence of other agents [7].

In a Single-Agent RL scenario, the agent interacts with a stationary environment, which is only changed by the agent's actions. In MARL, many agents influence the environment at the same time. This can result in non-stationarity of the environment as the state and reward of the individual agent depend on the behaviour of the other agents. The biggest problem of this non-stationarity is that it can lead to a violation of the Markov property, and thus the guarantee of convergence of many Single-Agent algorithm is no longer certain [7].

In addition to that, it is possible that the agents in a MARL want to achieve different learning goals that can even be opposing to each other [7]. These different goals can then lead to adversarial behaviour. Therefore, it is important to find the most common goal. Thus, we are searching for the best outcome for each agent, taking into account the most probable actions of all other agents.

4 Game Theory

For studying the behaviour of many agents in one system, game theoretical frameworks are helpful approaches, since game theory deals with the behaviour of many players following their own goals. In particular, game theory analyses strategies of competing participants mathematically, searching for the outcome of the player depending on the actions of all other players [4].

In many games, all player choose their action simultaneously. This concept leads to the definition of a normal-form game (NFG) as a triple $(N, \{S_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$, where $N = \{1, ..., n\}$ is the number of players, $\{S_1, S_2, ..., S_n\}$ are the sets of strategies for every player and $\{u_1, u_2, ..., u_n\}$ is a set of payoff functions [1]. The payoff functions define the reward that a player gets depending on his own action and the actions of all other players.

Based on the idea, that both players can choose their action arbitrary, the situation of both players can be analysed. Since every player wants the best payoff for itself, it is important to take the possibilities of the other players into account. Under the assumption that all player use a rational strategy to get the highest reward out of their situation, it is possible to know which actions the players will choose. This situation, in which the players will stick in the end, is called an equilibrium. The most common equilibrium is defined as follows [1]:

In the n-player normal-form game $(N, \{S_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$, the strategies $(s_1^*, s_2^*, \ldots, s_N^*)$ are a Nash equilibrium (NE) if, for each player i, s_i^* is player i's best response to the strategies specified for the $N-1$ other players $(s_1^*, ..., s_{i-1}^*, s_{i+1}^*, ..., s_N^*)$, i.e.

$$u_i(s_1^*, ..., s_N^*) \geq u_i(s_1^*, ..., s_{i-1}^*, s_i, s_{i+1}^*, ..., s_N^*)$$

for every feasible strategy s_i in S_i, that is, s_i^* solves

$$max_{s_i \in S_i} u_i(s_1^*, ..., s_{i-1}^*, s_i, s_{i+1}^*, ..., s_N^*).[1]$$

Calculating the NE leads to the strategy that is most likely to be played. However, it is important to know, that a NE does not represent the best outcome in total, since the strategies are dependent on the actions of the other players. This can be seen in famous games like the prisoner's dilemma [4].

4.1 Game Theory for MARL

Equivalent to the definition in game theory, we can apply the concepts to MARL, renaming the parameters. That means in MARL, the actions are similar to the strategies, while the reward is synonymous to the payoff. With this information, we can then define NFGs for MARL. A NFG in MARL is called a Markov Game [7] and is defined as a tuple $(N, S, \{A_i\}_{i=1}^n, P, \{R_i\}_{i=1}^n, \gamma)$, where $N = \{1, ..., n\}$ is the number of agents, S is the joint state space, $\{A_i\}_{i=1}^n$ is the action space for all agents, P is the transition probability, $\{R_i\}_{i=1}^n$ is the reward for all agents and γ is the discount factor [7]. As in NFG, the agents in Markov games choose their actions simultaneously.

5 MARL for Wireless Communication

In wireless communication, many devices want to communicate at the same time. Since resources for communication are limited, it is important to assign these resources precisely to the various devices. Thus, many devices compete for the same resources.

There are two types of communication in a communication cell, which are shown in Fig. 2. The most common way to communicate is through the base station (BS). This BS is also responsible for the distribution of resources within the communication cell. Thus, every device communicates with the BS to get their resources for communication. However, there is another way of communication, that allows the devices to communicate directly without interaction with the BS. This type of communication does not require the devices to communicate with the base station, thus saving resources [3].

If the devices communicate on their own, without regulation of the BS, the devices do not know which resources they can use. Thus, it is possible that some devices use the same resources for communication. This leads to an overlap of transmissions, which then prevent successful communication.

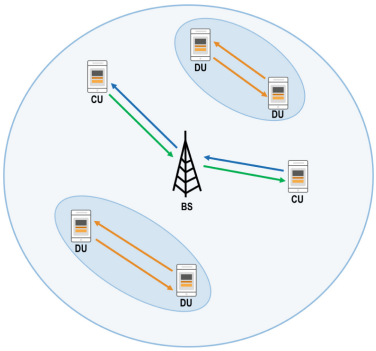


Fig. 2. In MARL the environment gives a state and a reward to all agents. The agents take their actions based on their state and thus, changing the environment.

To overcome this problem of lacking communication, we build a MARL scenario. In this scenario, the communicating devices represent the different agents. Every agents' goal is a reliable communication without disturbing the communication of other devices. A similar problem is shown in Zhi et al. [8], with some differences in the selection of the state and the RL method.

As an environment, we choose the communication cell with the BS. This environment gives a state to every agent that consists of the used frequency, the own Quality of Service (QoS) achieved by the last action, a vector of all neighbouring devices, and the frequency the neighbouring devices used in their last action. Given this state, all agents take their action. This action is the selection of a channel on that the agent wants to communicate in the next time step. Afterwards, the environment gives the new state to the agents, as well as the reward. Our reward is the sum of the QoS of all devices. This shared reward encourages the agent to a cooperative behaviour, since they have a shared learning goal.

5.1 Results

We set up a simulation of a communication cell as an RL environment with the above-mentioned restrictions for the states, actions and rewards. Then, we use a Q-Learning algorithm to let the agents learn a channel selection. We start with a two agent scenario. After a training time of 10.000 episodes, both agents have learned to choose their communication channel so that a reliable communication could happen. Figure 3 shows the Q-values of one agent. The Q-values describe how good the expected rewards are for different actions, depending on the actual state. The plot shows that the agent stays by his channel selection if it achieved a good reward, as can be seen in the upper right plot (agent stays by action 0) and the lower left plot (agent stays by action 1). If the agent achieves a bad reward in the last step, then it depends on the state if the agent chooses action 1 (as in the upper left plot) or action 0 (as in the lower right plot). Using these Q-values

as their strategy, both agents choose different channels for communication, what leads to a reliable communication.

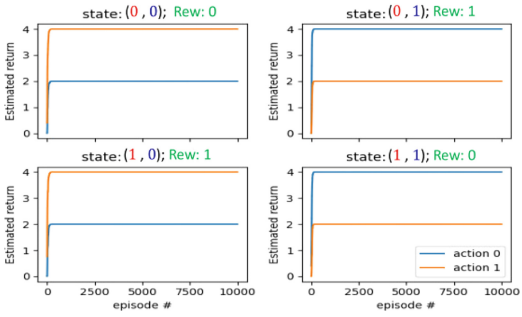


Fig. 3. Q-values of one agent in a two-player scenario for the four occurring states, showing that the agent chooses different actions depending on the state. The red channel selection of the state belongs to this agent, while the blue channel selection of the state belongs to the other agent.

After the agents in a two agent scenario have learned properly, we increase the number of agents, staying in the same scenario. Figure 4 shows the evaluation of the training runs with three and four agents.

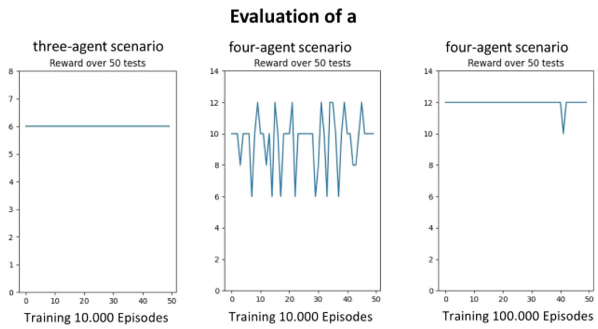


Fig. 4. Evaluation of trained strategies in a three- and a four-agent scenario, after a training of 10.000 episodes and 100.000 episodes. The maximum reward is 6 for three agents and 12 for four agents.

In the three-agent scenario, the agents have learned an optimal strategy to achieve reliable communication without disturbing the communication of other agents after a training time of 10.000 episodes. This can be seen in the left plot of Fig. 4, since all test scenarios there are successfully completed by the agents with the maximum reward of 6. This evaluation happens in the simulation of

a communication cell in which the agents are confronted with the previously trained task of correctly selecting their own channels. It is designed in such a way that all possible scenarios that can occur in the communication cell are tested in the 50 test scenarios.

The right side of the plot in Fig. 4 shows the evaluation results of the four-agent scenario. The agents are not trained well after 10.000 episodes. This can be seen by the few testing scenarios that achieve the maximum reward of 12. After a longer training run of 100.000 episodes, the agents behave much better. Now they get the maximum reward of 12 nearly in all testing scenarios. However, there is still one testing scenario without maximum reward, which means, that the agents have not learned the optimal strategy yet.

6 Conclusion and Future Work

MARL is a difficult sub-field of RL, where many challenges lead to difficult training assumptions. Despite these difficult conditions, it is important to continue research in MARL, because this area of methods can be used to represent and solve more complicated real-world problems.

In the future, it is important to consider other methods for our wireless communication problem. There are for example new algorithms that combine RL for Multi-Agents with Game Theory, like NashQ.

Moreover, we want to increase the number of agents in the environment, since more than four devices can occur in real-world communication. Another addition is the possibility that the devices move, as communication cells consist of many devices that enter and leave the environment and move around in it, like people with their mobile phones.

References

1. Gibbons, R.: A Primer in Game Theory. Harvester Wheatsheaf, New York (1992)
2. Jagtap, S., Thomas, A., Gadiwan, S., Mishra, S.: Multi-agent reinforcement learning-implementation of hide and seek. In: 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT) (2022)
3. Molisch, A.F.: Wireless Communications. Wiley Publishing (2011)
4. Osborne, M.J.: An Introduction to Game Theory. Oxford University Press, New York (2004)
5. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press (2018)
6. Zhang, H., Zhang, S.: Multi-agent reinforcement learning. In: Deep Reinforcement Learning: Fundamentals. Research and Applications, pp. 335–346. Springer, Singapore (2020)
7. Zhang, K., Yang, Z., Başar, T.: Multi-agent reinforcement learning: a selective overview of theories and algorithms. In: Handbook of Reinforcement Learning and Control, pp. 321–384. Springer (2021)
8. Zhi, Y., Tian, J., Deng, X., Qiao, J., Lu, D.: Deep reinforcement learning-based resource allocation for D2D communications in heterogeneous cellular networks. Digit. Commun. Netw. **8**(5), 834–842 (2022)

Order Determination in Second-Order Source Separation Models Using Data Augmentation

Una Radojičić[1] and Klaus Nordhausen[2(✉)]

[1] Vienna University of Technology, 1040 Vienna, Austria
una.radojicic@tuwien.ac.at
[2] University of Jyväskylä, 40014 Jyväskylä, Finland
klaus.k.nordhausen@jyu.fi

Abstract. We propose a robust estimator for the number of latent components in an internal noise model within the second-order source separation (SOS) framework. Our approach utilizes a data augmentation strategy in conjunction with the robust SOS approach eSAM-AMUSE, which combines information from eigenvalues and variations of eigenvectors of eSAM-AMUSE. The resulting dimension estimate can be visualized using a ladle plot. Through a simulation study, we demonstrate the superior properties of the new estimator, which outperforms the bootstrap-based AMUSEladle estimator.

Keywords: AMUSE · Blind source separation · Robustness · Spatial signs

1 Introduction

Multivariate time series are inherently difficult to model, leading to the adoption of various simplifying structural assumptions to make the modeling more tractable. One popular approach is the second-order source separation (SOS) model

$$\boldsymbol{x}_t = \boldsymbol{\Omega}\boldsymbol{z}_t + \boldsymbol{\mu}, \tag{1}$$

where $\boldsymbol{x}_t$ is the p-variate observable time series and $\boldsymbol{z}_t$ consists of p standardized completely uncorrelated second order stationary time series. The full rank matrix $\boldsymbol{\Omega} \in \mathbb{R}^{p\times p}$ is called the mixing matrix and $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes the location of $\boldsymbol{x}_t$. SOS thus constitutes a specific case of blind source separation (BSS), aiming to recover $\boldsymbol{z}_t$ by identifying an unmixing matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p}$ such that $\boldsymbol{\Gamma}(\mathbf{x}_t - \boldsymbol{\mu}) = \boldsymbol{z}_t$. For more details about BSS and other BSS time series models see for example [2,9,11]. The central concept behind SOS/BSS is that instead of directly modeling the original p-variate process, the p processes within $\boldsymbol{z}_t$ can be modeled individually. However, this approach may still prove cumbersome when

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 371–379, 2024.
https://doi.org/10.1007/978-3-031-65993-5_46

p is large, prompting another common assumption, namely that $\boldsymbol{x}_t$ is driven by $q < p$ signal processes and is distorted by noise. To accommodate the lower signal dimension and noise into BSS/SOS [11] discuss, among others, the internal noise model (INM)

$$\boldsymbol{x}_t = \boldsymbol{\Omega}\boldsymbol{z}_t + \boldsymbol{\mu} = \boldsymbol{\Omega}_1\boldsymbol{z}_{1,t} + \boldsymbol{\Omega}_2\boldsymbol{z}_{2,t} + \boldsymbol{\mu}, \tag{2}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \in \mathbb{R}^{p\times p}$ is non-singular, $\boldsymbol{\Omega}_1 \in \mathbb{R}^{p\times q}$, $\boldsymbol{\Omega}_2 \in \mathbb{R}^{p\times(p-q)}$, and $\boldsymbol{\mu} \in \mathbb{R}^p$. The p-variate latent process $\boldsymbol{z}_t = (\boldsymbol{z}_{1,t}^\top, \boldsymbol{z}_{2,t}^\top)^\top$, can here be decomposed into two parts, the q-variate signal $\boldsymbol{z}_{1,t}$ and the $(p-q)$-variate noise $\boldsymbol{z}_{2,t}$. In INM, the assumption is typically made that q is unknown and requires estimation, a process referred to as order determination. Hypothesis tests for q are discussed in [6,15]. Additionally, a computationally intensive estimator, the ladle, based on time-series bootstrapping, is introduced in [10]. It is noteworthy that these estimation methods, while widely used, share a common limitation - they are moment-based, making them inefficient in scenarios with heavy-tailed signals and lacking robustness to outliers.

The objective of this paper is to introduce a novel estimator for q in the INM model that is less computationally demanding while simultaneously being more robust. To achieve this, we provide further insights into Model (2), discuss the estimation of $\boldsymbol{\Omega}$ within this model, and subsequently introduce our estimator, which is illustrated through a simulation study.

2 Second-Order Source Separation Internal Noise Model

In internal noise models, under certain assumptions, the signals can be consistently estimated. In particular in the SOS INM model (2), a popular and easy-to-compute approach to recover the signals is called AMUSE [8,14], for which $\boldsymbol{z}_{1,t}$ is identifiable up to the order of its components and their signs under the following assumptions:

- **(A1)** $E(\boldsymbol{z}_t) = \boldsymbol{0}$ and $\mathrm{COV}(\boldsymbol{z}_t) = \boldsymbol{I}_p$;
- **(A2)** $\boldsymbol{z}_{1,t} \in \mathbb{R}^q$ and $\boldsymbol{z}_{2,t} \in \mathbb{R}^{p-q}$ are mutually independent;
- **(A3)** The components of $\boldsymbol{z}_{1,t}$ are uncorrelated and exhibit second-order serial dependence of the form $\mathrm{ACOV}_\tau(\boldsymbol{z}_t) = \boldsymbol{D}_\tau$ for all τ, where ACOV_τ denotes the autocovariance matrix at lag τ, and $\boldsymbol{D}_\tau$ is a diagonal matrix;
- **(A4)** The components of $\boldsymbol{z}_{2,t}$ are white noise;
- **(A5)** $\mathrm{ACOV}_\tau(\boldsymbol{z}_{1,t})$ has distinct non-zero diagonal values for the chosen lag τ.

For simplicity, considering first the case with $p = q$, AMUSE begins by whitening $\boldsymbol{x}_t$ using the covariance matrix, i.e. $\boldsymbol{x}_t^{st} := \mathrm{COV}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu})$. The whitened process has therefore $\mathrm{COV}(\boldsymbol{x}_t^{st}) = \boldsymbol{I}_p$, meaning we still have the freedom to choose an additional rotation $\boldsymbol{U}$ to recover the latent signals, i.e., $\boldsymbol{z}_t = \boldsymbol{U}^\top \boldsymbol{x}_t^{st}$. AMUSE obtains this additional rotation $\boldsymbol{U} = \boldsymbol{U}_\tau$ by diagonalizing $\mathrm{ACOV}_\tau(\boldsymbol{x}_t^{st})$, the autocovariance matrix of the standardized series at the chosen lag $\tau \geq 1$, using the eigendecomposition $\mathrm{ACOV}_\tau(\boldsymbol{x}_t^{st}) = \boldsymbol{U}_\tau \boldsymbol{D}_\tau \boldsymbol{U}_\tau{}^\top$. The AMUSE unmixing matrix is then $\boldsymbol{\Gamma}_\tau = \mathrm{COV}(\boldsymbol{x}_t)^{-1/2}\mathbf{U}_\tau$, and the latent components are found as $\boldsymbol{\Gamma}_\tau(\boldsymbol{x}_t - \boldsymbol{\mu})$.

Now, if $q < p$, then all noise components exhibit zero eigenvalues in $\text{ACOV}_\tau(\boldsymbol{x}_t^{st})$. Consequently, the signals can be reconstructed using only the eigenvectors corresponding to non-zero eigenvalues.

AMUSE has however two drawbacks. Firstly, its performance depends a lot on the selection of τ and suffers from heavy tail signals or outliers. Essentially, the lag should be selected such that the signal eigenvalues are as distinct as possible, and visual analytic tools for this purpose are for example discussed in [12]. An extension of AMUSE that, to some extent mitigates this limitation is called SOBI and jointly diagonalizes several autocovariance matrices. SOBI is however beyond the scope of this paper and we refer the reader to [1,7].

Secondly, AMUSE suffers from a lack of robustness against heavy tail signals or outliers. Namely, in AMUSE, all population quantities needed (location, covariance, and autocovariance matrix) are in practice replaced by their moment-based finite sample estimators. Following [4] we advocate instead *affine equivariant sign autocovariance AMUSE* (eSAM-AMUSE) which robustifies AMUSE by replacing mean and covariance matrix with the corresponding spatial sign-based affine equivariant Hettmansperger-Randles estimators of location and scatter [3]. In the same spirit [4] substitute ACOV_τ by the spatial-sign autocovariance matrix, which for centered data is defined as $\boldsymbol{R}_\tau(\boldsymbol{x}_t) = E\left(\frac{\boldsymbol{x}_t \boldsymbol{x}_{t+\tau}^\top}{\|\boldsymbol{x}_t\|\|\boldsymbol{x}_{t+\tau}\|}\right)$.

For a realization $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ from an INM model the finite sample version of eSAM-AMUSE can be obtained as

eSAM 1 Robust standardization of the data $\boldsymbol{x}_t^{st} = \hat{\boldsymbol{S}}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \hat{\boldsymbol{T}}(\boldsymbol{x}_t))$, where $\hat{\boldsymbol{T}}(\boldsymbol{x}_t))$ and $\hat{\boldsymbol{S}}(\boldsymbol{x}_t))$ are the Hettmansperger-Randles estimates of location and scatter, respectively;

eSAM 2 Robust autocovariance estimation based on the spatial-sign autocovariance matrix:

$$\hat{\boldsymbol{R}}_\tau(\boldsymbol{x}_t^{st}) = \frac{1}{T-\tau}\sum_{i=1}^{T-\tau}\left(\frac{\boldsymbol{x}_t^{st}\boldsymbol{x}_{t+\tau}^{st\,\top}}{\|\boldsymbol{x}_t^{st}\|\|\boldsymbol{x}_{t+\tau}^{st}\|}\right)$$

eSAM 3 Symmetrizing the robust autocovariance: $\boldsymbol{S}_\tau = \frac{1}{2}\left(\hat{\boldsymbol{R}}_\tau + \hat{\boldsymbol{R}}_\tau{}^\top\right)$ to ensure that the estimate is symmetric as the population version;

eSAM 4 Computing the eigendecomposition $\boldsymbol{S}_\tau = \mathbf{U}_\tau \boldsymbol{\Lambda}_\tau \mathbf{U}_\tau^\top$. For later convenience we assume that the eigenvalues in $\boldsymbol{\Lambda}_\tau$ are ordered so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_p| \geq 0$.

Note that for eSAM-AMUSE to be able to recover the latent signals we need to replace assumption (A5) with:

(A5*) $\boldsymbol{z}_t$ is a symmetric process and $\boldsymbol{R}_\tau(\boldsymbol{z}_{1,t})$ has distinct non-zero diagonal values for the chosen lag τ.

In the INM model, the signals are linked to components with non-zero eigenvalues in $\boldsymbol{\Lambda}_\tau$, while noise components correspond to those with zero eigenvalues. Therefore, order determination involves deciding which eigenvalues are zero. This

paper employs the "augmentation" procedure from [5] for order determination. This procedure combines information from both the eigenvectors and the eigenvalues of the autocovariance matrices, providing a comprehensive approach to the task.

3 Augmentation Procedure

For simplicity and clarity, we start by motivating the data augmentation for AMUSE, on the population level.

3.1 The Main Idea

Let $\boldsymbol{x}_t^{st} = \mathrm{COV}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu})$, as discussed in Sect. 2. For fixed lag τ, let $\mathrm{ACOV}_\tau(\boldsymbol{x}_t^{st}) = \boldsymbol{\Sigma}_\tau$ be the autocovariance matrix of $\boldsymbol{x}_t^{st}$, and let $\mathbf{U}_\tau$ be the corresponding matrix of eigenvectors. Then, $\boldsymbol{\Gamma}_\tau$ is, up to sign changes, guaranteed to recover latent components and $\mathrm{rank}(\boldsymbol{\Sigma}_\tau)$ corresponds to the number of latent components that exhibit the serial dependence at the lag τ. Especially, since $\boldsymbol{z}_{2,t}$ is assumed to be white noise, $\mathrm{rank}(\boldsymbol{\Sigma}_\tau) = d \leq q$, where $d = q$ in the case where all the signal components exhibit serial dependence at lag τ. Thus, the problem of separating the signal from the noise involves estimating the rank of $\boldsymbol{\Sigma}_\tau$.

For fixed $r \geq 1$, let now $\boldsymbol{\varepsilon}_t \in \mathbb{R}^r$, to be a white noise time series independent of $\boldsymbol{x}_t$, with $\mathrm{COV}(\boldsymbol{\varepsilon}_t) = \mathbf{I}_r$, and define the *augmented standardized time-series*

$$\boldsymbol{x}_t^{st*} = (\boldsymbol{x}_t^{st\top}, \boldsymbol{\varepsilon}_t^\top)^\top \in \mathbb{R}^{p+r}.$$

Then, assuming all the signal components $\mathbf{z}_{t,1}$ exhibit serial dependence at the lag τ, the autocovariance function of $\boldsymbol{x}_t^{st*}$ at lag τ is

$$\mathrm{ACOV}(\boldsymbol{x}_t^{st*}) := \boldsymbol{\Sigma}_\tau^* = \begin{pmatrix} \boldsymbol{\Sigma}_\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_\tau \boldsymbol{\Lambda}_\tau \mathbf{U}_\tau^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathrm{rank}(\boldsymbol{\Sigma}_\tau^*) = \mathrm{rank}(\boldsymbol{\Sigma}_\tau) = q$. Let now $\boldsymbol{u}_i^* = (\boldsymbol{u}_i^\top, \boldsymbol{u}_i^{\mathrm{AUG}\top})^\top$, $\boldsymbol{u}_i \in \mathbb{R}^p$, $\boldsymbol{u}_i^{\mathrm{AUG}} \in \mathbb{R}^r$ be the ith eigenvector of $\boldsymbol{\Sigma}_\tau^*$, belonging to the eigenvalue λ_i^*, where the ordering of the eigenvalues is such that $|\lambda_1^*| \geq |\lambda_2^*| \geq \cdots \geq |\lambda_p^*|$. Then,

$$\boldsymbol{\Sigma}_\tau^* \boldsymbol{u}_i^* = ((\boldsymbol{\Sigma}_\tau \boldsymbol{u}_i)^\top, \mathbf{0}^\top)^\top = \lambda_i^* \boldsymbol{u}_i^* = (\lambda_i^* \boldsymbol{u}_i{}^\top, \lambda_i^* \boldsymbol{u}_i^{\mathrm{AUG}\top})^\top, \tag{3}$$

where the penultimate equality holds since $\boldsymbol{u}_i^{\mathrm{AUG}} \in \mathbb{R}^r$ is the eigenvector of $\boldsymbol{\Sigma}_\tau^*$. Specifically, (3) implies that $\lambda_i^* \boldsymbol{u}_i^{\mathrm{AUG}} = \mathbf{0}$, further giving that $|\lambda_i^*| \|\boldsymbol{u}_i^{\mathrm{AUG}}\| = 0$. For $i \leq q$, $|\lambda_i^*| = |\lambda_i| > 0$, implying that $\|\boldsymbol{u}_i^{\mathrm{AUG}}\|^2 = 0$. The same does not hold for the latter eigenvectors, i.e. for $i > q$. Thus, examining the norms of the latter, the "augmented" part of the eigenvectors of $\boldsymbol{\Sigma}_\tau{}^*$, one can conclude the $\mathrm{rank}(\boldsymbol{\Sigma}_\tau) = q$.

3.2 Augmentation for ESAM-AMUSE

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T \in \mathbb{R}^p$ be a realization from Model (2), and let $\boldsymbol{x}_t^{st} = \hat{\boldsymbol{S}}(\boldsymbol{x}_t)^{-1/2}(\boldsymbol{x}_t - \hat{\boldsymbol{T}}(\boldsymbol{x}_t))$, $t = 1, \ldots, T$ be its robust standardization; see (eSAM 1). For fixed $r \geq 1$, let $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T \in \mathbb{R}^r$, be a realization of a white noise time series with zero mean and identity covariance, independent of $\boldsymbol{x}_t$. We augment now the standardized time series with the simulated noise to obtain

$$\boldsymbol{x}_t^{st*} = (\boldsymbol{x}_t^{st\top}, \boldsymbol{\varepsilon}_t^{\top})^{\top}, \quad i = 1, \ldots, T. \tag{4}$$

Remark 1. In Sect. 3.1, the standardization of the time series preceded the augmentation. On the population level, as $\boldsymbol{x}_t$ and $\boldsymbol{\varepsilon}_t$ are independent and the noise was assumed to be standardized, these two operations are interchangeable. On the sample level, we choose to first standardize $\boldsymbol{x}_t$ only, thus avoiding the estimation and inversion of the covariance of the $p + r$-dimensional augmented data, as its dimension can become rather large if r is chosen large (especially in comparison with the length T).

Recall $\hat{\boldsymbol{S}}_\tau^* = \frac{1}{2}\left(\hat{\boldsymbol{R}}_\tau^* + \hat{\boldsymbol{R}}_\tau^{*\top}\right)$, the robust symmetrized estimate of the lagged spatial sign autocovariance matrix of $\boldsymbol{x}_t^{st*}$ as introduced in (eSAM 3). Let further $\mathbf{U}_\tau^* = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p]$ be the matrix of eigenvectors of $\hat{\boldsymbol{S}}_\tau^*$, and let $\boldsymbol{\Lambda}_\tau^* = \operatorname{diag}\left(\mathbf{U}_\tau^{*\top}\hat{\boldsymbol{S}}_\tau^*\mathbf{U}_\tau^*\right)^2 \in \mathbb{R}^{p\times p}$ with the absolute eigenvalues of $\hat{\boldsymbol{S}}_\tau^*$ in decreasing order on the diagonal (as demonstrated in (eSAM 4)).

To reduce the effect of randomness in the augmentation, the augmentation procedure is conducted independently $s \geq 1$ times, and we compute the matrix of eigenvectors $\mathbf{U}_\tau^j$ of $\hat{\boldsymbol{S}}_\tau^{*j}$ for each replicate $j = 1, \ldots, s$. Additionally, for $j = 1, \ldots, s$, we denote by $\boldsymbol{u}_i^{\mathrm{AUG},j} \in \mathbb{R}^r$ the latter (augmented) part of the ith column of the matrix $\mathbf{U}_\tau^j$ in the jth replicate. Taking $\boldsymbol{u}_0^{\mathrm{AUG},j} := \mathbf{0}$, the eigenvector information about the rank of $\boldsymbol{\Sigma}_\tau$ is captured by the function

$$f : \{0, 1, \ldots, p\} \to \mathbb{R}, \quad f(i) = \frac{1}{s}\sum_{j=1}^{s} \|\boldsymbol{u}_i^{\mathrm{AUG},j}\|^2.$$

Scree Plot for AMUSE Unlike in PCA, the ordering of the eigenvalues of $\boldsymbol{\Sigma}_\tau$, in principle, does not tell us anything about the component's relevance and we are only interested in how many are non-zero. To estimate the rank of $\boldsymbol{\Sigma}_\tau$ based on this property we can however still consider the *scree plot* of the absolute values of the eigenvalues.

Let $\lambda_1, \ldots, \lambda_p$ be the diagonal elements of $\boldsymbol{\Lambda}_\tau$ ordered as described in (eSAM-4). Assuming that all the signal components exhibit serial dependence at the lag τ, we expect the first q eigenvalues in $\boldsymbol{\Lambda}_\tau$ to be non-zero, while the latter ones be small, reflecting the white-noise nature of the noise. Setting $\lambda_{p+1} := 0$ and mimicking [5], we define the normalized scree plot curve as

$$\Phi : \{0, 1, \ldots, p\} \to \mathbb{R}, \quad \Phi(l) = |\lambda_{l+1}| / \left(\sum_{i=1}^{l+1} |\lambda_i|\right).$$

Unlike in [5], we drop the addition of 1 in the denominator to make the method scale invariant. Such property is particularly important for robust methods, in which the scale of the (auto)covariance is often miss-estimated. In an extreme noise case ($\lambda_i \approx 0$, $\forall i$), to improve the numerical stability, one is advised to add a small constant $\delta > 0$ to the denominator.

Augmentation Estimator The augmentation estimator, as in [5], is obtained by summing the eigenvalue function Φ and the eigenvector function f, i.e.,

$$g : \{0, 1, \ldots, p\} \to \mathbb{R}, \quad g(k) = \Phi(k) + \sum_{i=1}^{k} f(i),$$

whose minimizer is then taken to be the augmentation-based estimate of the number of signal components, i.e.,

$$\hat{q} = \underset{k \in \{0,\ldots,p-1\}}{\arg\min} \; g(k). \tag{5}$$

Heuristically, the procedure is reasonable, as for $k < q$ (and $q > 0$), the eigenvalue component $\Phi(k)$ in (5) tends to be large, while the eigenvector component $f(k)$ is small. Conversely, for $k > q$, the situation is reversed, with the eigenvalue part being small and the eigenvector part being large. At the correct dimension $k = q$, both components are close to 0, indicating that the sum curve g is (on a population level) precisely minimized at $k = q$. Additionally, in the scenario of extreme noise where $q = 0$, the eigenvalue part in (5) is always negligible, and the eigenvector part is consistently large, except for the case $k = 0$ where it vanishes, again leading to the minimum occurring at $k = 0 = q$.

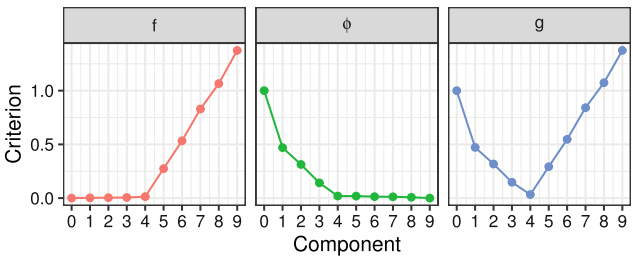


Fig. 1. Ladle plot g, based on an augmented eSAM-AMUSE estimate using $r = s = 10$ and $\tau = 1$ together with its individual parts. The estimated signal dimension is $\hat{q} = 4$

Figure 1 depicts the plots of Φ, f, and g, demonstrating the described behavior for one of the Gaussian data sets used later in the simulation study. Reference [5] show that in many multivariate iid models, under suitable assumptions, the procedure yields a consistent estimator of the latent dimension, while [13] show that it yields a consistent estimator in the tensorial additive noise model, in the context of tensor PCA. Exploring the validity of that claim for our scenario

involving dependent time series data and eSAM-AMUSE is an avenue for future research. In this work, we, however, focus on demonstrating the validity of the proposed approach in a simulation study.

4 Simulations

To evaluate order determination with the augmented estimator, we conducted a simulation study initially involving four independent Gaussian signals generated from autoregressive moving average (ARMA) processes: $\text{ARMA}_{(0.4,0.2);(0.3,0.1)}$, $\text{ARMA}_{(0.6,0.2,0.1);0}$, $\text{ARMA}_{0.3;(0.7,0.2)}$, and $\text{ARMA}_{0;(0.2,0.6,0.4,0.2)}$. Five Gaussian white noise processes were added, and the nine latent components were mixed at each repetition with a random mixing matrix $\mathbf{\Omega}$. This process was repeated 2000 times, generating a time series of length $T = 2000$. To evaluate the impact of outliers and heavy tails, we modified the Gaussian setting in three ways.

In the isolated outlier setting, 15 observations, each 100 time points apart, were independently contaminated with random values of the form wa, where $w \in \{-1, 1\}$ with equal probability and $a \sim \mathcal{N}(20, 1)$. In the patchy outlier setting, 15 successive observations were contaminated independently in each component with values from $\mathcal{N}(20, 1)$. In the heavy tails setting, Gaussian innovations in the signals were replaced with innovations from a 4-variate spherical t_3 distribution scaled to have an identity covariance matrix.

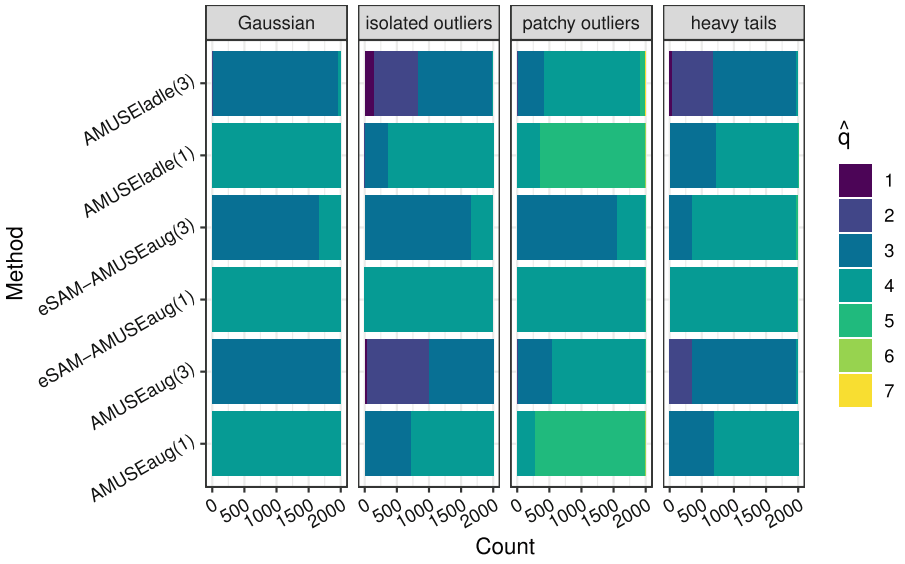


Fig. 2. Counts of the estimated signal dimension in the simulation study based on 2000 repetitions. The true signal dimension is $q = 4$

For augmented estimators, we consistently used $s = 10$ and $r = 10$. As a reference, we employed the AMUSE ladle estimator by [10] with 200 stationary

block bootstrap samples and an expected block length of 20. Denoting the three estimators as AMUSEaug(τ), eSAM-AMUSEaug(τ), and AMUSEladle(τ), we compared the estimation of q using $\tau = 1$ and $\tau = 3$. The resulting estimated $\hat{q}$ for all methods across the 2000 repetitions are visualized in Fig. 2. The results clearly show that the lag choice is crucial and $\tau = 1$ performs best. While all three methods work very well in the Gaussian case it is clear that only eSAM-AMUSEaug(1) can cope in both outlier settings and with heavy tailed signals.

5 Conclusion

Most BSS applications exhibit serial dependence, and the primary objective is typically to model the data univariately and reduce the dimensionality of the data while minimizing the impact of outliers. For the later, order determining is crucial, and in this paper, we demonstrate how the augmentation strategy proposed by [5] can be extended to dependent data within the context of the INM model in an SOS framework using the eSAM-AMUSE estimator. However, achieving excellent performance hinges on the choice of lag for the eSAM-AMUSE method. Therefore, future work will aim to extend the current approach to SOBI-type estimators.

Acknowledgments. The work of Una Radojičić is supported by the Austrian Science Foundation, project number I 5799-N.

References

1. Belouchrani, A., Abed Meraim, K., Cardoso, J.F., Moulines, E.: A blind source separation technique based on second order statistics. IEEE Trans. Signal Process. **45**, 434–444 (1997)
2. Comon, P., Jutten, C.: Handbook of Blind Source Separation. Independent Component Analysis and Applications. Academic Press (2010)
3. Hettmansperger, T.P., Randles, R.H.: A practical affine equivariant multivariate median. Biometrika **89**(4), 851–860 (2002)
4. Ilmonen, P., Nordhausen, K., Oja, H., Theis, F.: An affine equivariant robust second-order BSS method. In: Vincent, E. et al. (Eds.) Latent Variable Analysis and Signal Separation, pp. 328–335. Springer (2015)
5. Luo, W., Li, B.: On order determination by predictor augmentation. Biometrika **108**, 557–574 (2021)
6. Matilainen, M., Nordhausen, K., Virta, J.: On the number of signals in multivariate time series. In: Deville, Y. et al. (Eds.) Latent Variable Analysis and Signal Separation, pp. 248–258. Springer (2018)
7. Miettinen, J., Illner, K., Nordhausen, K., Oja, H., Taskinen, S., Theis, F.: Separation of uncorrelated stationary time series using autocovariance matrices. J. Time Ser. Anal. **37**, 337–354 (2016)
8. Miettinen, J., Nordhausen, K., Oja, H., Taskinen, S.: Statistical properties of a blind source separation estimator for stationary time series. Statist. Probab. Lett. **82**, 1865–1873 (2012)

9. Nordhausen, K., Oja, H.: Independent component analysis: a statistical perspective. WIREs Comp. Stats. **10**, e1440 (2018)
10. Nordhausen, K., Virta, J.: Ladle estimator for time series signal dimension. In: 2018 IEEE Statistical Signal Processing Workshop (SSP), pp. 428–432 (2018)
11. Pan, Y., Matilainen, M., Taskinen, S., Nordhausen, K.: A Review of Second-Order Blind Identification Methods.WIREs Comp. Stats. 14, e1550 (2022)
12. Piccolotto, N., Bögl, M., Gschwandtner, T., Muehlmann, C., Nordhausen, K., Filzmoser, P., Miksch, S.: TBSSvis: visual analytics for temporal blind source separation. Vis. Inform. **6**(4), 51–66 (2022)
13. Radojicic, U., Lietzen, N., Nordhausen, K., Virta, J.: Order determination for tensor-valued observations using data augmentation. Preprint available at [arXiv2207.10423] (2022)
14. Tong, L., Soon, V., Huang, Y., Liu, R.: AMUSE: a new blind identification algorithm. In: Proceedings of IEEE International Symposium on Circuits and Systems, pp. 1784–1787 (1990)
15. Virta, J., Nordhausen, K.: Determining the signal dimension in second order source separation. Statist. Sinica **31**, 135–156 (2021)

Check for updates

Clustering Ordinal Data Via Parsimonious Models

Monia Ranalli and Roberto Rocci(✉)

Sapienza University of Rome, Rome, Italy
{monia.ranalli,roberto.rocci}@uniroma1.it

Abstract. This review presents some parsimonious models to cluster two-way and three-way ordinal data. They are formulated has a reparameterization of a finite mixture of Gaussians that is partially observed through a discretization of its variates. Model parameters are estimated using a composite likelihood approach in order to reduce the numerical complexity. The parsimony is obtained by reducing the dimensionality of the variable's space within and/or between the components.

Keywords: Model-based clustering · Parsimonious modeling · Composite likelihood · Ordinal data

1 Introduction

A frequently used and very well known model in cluster analysis is the finite mixture of Gaussians [11]. It is based on the assumption that a random vector (r.v.) $\mathbf{y}$ has a probability density function (p.d.f.) which can be written as

$$f(\mathbf{y};\vartheta) = \sum_{g=1}^{G} \phi_J(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)p_g, \tag{1}$$

where $\phi_J(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)$ is the p.d.f. of a J-variate Gaussian random vector with mean $\boldsymbol{\mu}_g$ and covariance $\boldsymbol{\Sigma}_g$. The idea is to interpret the g-th Gaussian component as a cluster in the population with weight p_g $(g = 1, 2, \ldots, G)$. Following this model, an observation $\mathbf{y}$ is classified into the cluster l if

$$l = \text{argmax}_g p(g|\mathbf{y}) = \text{argmax}_g \frac{\phi_J(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)p_g}{\sum_{h=1}^{G} \phi_J(\mathbf{y};\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_h)p_h}, \tag{2}$$

where $p(g|\mathbf{y})$ is the posterior probability that the observation $\mathbf{y}$ comes from the component g.

Such a model works quite well in practice but it is not appropriate when the observed variables are ordinal. In order to still use the finite mixture of Gaussians, some practitioners transform each variable as numeric by computing a score for each level of the ordinal scale. This makes the computation feasible

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 380–387, 2024.
https://doi.org/10.1007/978-3-031-65993-5_47

but it does not make the variable quantitative nor normally distributed into the clusters (see [13] and references herein for a discussion). In the literature there are mainly two approaches to face with ordinal data: 1) Item Response Theory (IRT) [2]; 2) Underlying Response Variable (URV) [12]. The first approach is based on the hypothesis of "local independence", where the observed variables are independent given some latent variables. The clustering structure is specified through some latent variables following, for example, a discrete distribution [1] or a finite mixture of Gaussians [4]. On the other hand, in the URV approach each observed variable is assumed to be a discretization of a continuous latent variable. Even in this case the clustering structure is specified through the latent variables, for example assuming that they are distributed like a finite mixture of Gaussians [13]. The two approaches are not very different. For example, Takane and de Leeuw have shown how they can be equivalent in the case of the Factor Analysis model for ordinal data [18]. However, here we prefer the URV approach because it leads to models more easily comparable with the ones proposed for continuous data. This allows us to draw on the vast existing literature for continuous data in order to propose and understand new models.

In particular, here we consider the unsupervised classification model where the ordinal variables are generated by underlying continuous variables, distributed like a finite mixture of Gaussians, which are partially observed through their ordinal counterparts. The use and interpretation of this model is quite simple. However, in practice we have two main, strongly related, problems: (1) the computation of the maximum likelihood estimates can be rather cumbersome as the number of ordinal variables increases; (2) the number of parameters can be very large. In this review, we focus the attention on the second problem. First, in Sect. 2, we introduce the general model that can be used to cluster ordinal data and explain how the computational burden can be drastically reduced by adopting a composite likelihood approach for the estimation of model parameters. Then, three different parsimonious reparameterizations are presented for data organised in two-way (Sects. 3 and 4) and three-way (Sect. 5) form. They are all based on the dimensionality reduction of the variable's space: within the components (Sect. 3), in the total space (Sect. 4) or between the components (Sect. 5). A comparison among the three reparameterizations, along with some final comments and remarks, closes the paper in Sect. 6.

2 General Model

Let $\mathbf{x} = [x_1, x_2, \ldots, x_J]'$ be a r.v. where x_j is an ordinal variable with C_j categories for $j = 1, 2, \ldots, J$. There are $R = \prod_{j=1}^{J} C_j$ possible response patterns of the form $\mathbf{c}_r = [c_1, c_2, \ldots, c_J]'$ where $c_j \in \{1, 2, \ldots, C_j\}$ is the category of the j-th variable. Let $\mathbf{y}$ be a latent r.v. distributed like the heteroscedastic Gaussian mixture (1). Under the URV approach, the ordinal variables $\mathbf{x}$ are considered as the following discretization of $\mathbf{y}$

$$\gamma_{c_j-1}^{(j)} \leq y_j < \gamma_{c_j}^{(j)} \Leftrightarrow x_j = c_j,$$

where $-\infty = \gamma_0^{(j)} < \gamma_1^{(j)} < \ldots < \gamma_{C_j-1}^{(j)} < \gamma_{C_j}^{(j)} = +\infty$ are the thresholds defining the C_j categories of variable x_j. The probability of the response pattern $\mathbf{c}_r$ is given by

$$\begin{aligned}\Pr(\mathbf{c}_r;\vartheta) &= \Pr(x_1 = c_1, \ldots, x_J = c_J;\vartheta) \\ &= \sum_{g=1}^{G} p_g \int_{\gamma_{c_1-1}^{(1)}}^{\gamma_{c_1}^{(1)}} \cdots \int_{\gamma_{c_J-1}^{(J)}}^{\gamma_{c_J}^{(J)}} \phi_J(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)d\mathbf{y} \\ &= \sum_{g=1}^{G} p_g \pi_r(\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g,\gamma),\end{aligned}$$

where $\pi_r\left(\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g,\boldsymbol{\gamma}\right)$ is the probability of response pattern $\mathbf{c}_r$ in the cluster g, $\boldsymbol{\gamma}$ is the set of thresholds and ϑ is the set of parameters. For a random i.i.d. sample of size n the log-likelihood is

$$\ell(\vartheta;\mathbf{x}) = \sum_{r=1}^{R} n_r \log\left[\sum_{g=1}^{G} p_g \pi_r\left(\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g,\boldsymbol{\gamma}\right)\right], \tag{3}$$

where n_r is the sample frequency of response pattern $\mathbf{c}_r$ and $\sum_{r=1}^{R} n_r = n$. The evaluation of this log-likelihood could be rather burdensome because it implies the calculation of several multidimensional integrals. Ranalli and Rocci [13] proposed to solve this problem by adopting a composite likelihood approach [9] where the parameters are estimated by maximizing a surrogate function, named composite likelihood. Such a function is the product of some sub-likelihoods obtained by considering only some marginal and/or conditional distributions instead of the complete joint probability mass function. In particular, Ranalli and Rocci proposed to build the surrogate function by considering all the bivariate marginals. This approach leads to computationally efficient estimators that, under some regularity conditions, in several cases are only slightly less efficient than the corresponding maximum likelihood counterparts [20].

3 Within Dimensionality Reduction

In this approach the number of parameters is reduced by assuming that $\mathbf{y}$ follows a finite mixture of factor analyzers [6,10]

$$f(\mathbf{y}) = \sum_{g=1}^{G} p_g \phi_J(\boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g) \tag{4}$$

where $\boldsymbol{\Lambda}_g$ is the $J \times Q(< J)$ matrix of factor loadings, and $\boldsymbol{\Psi}_g$ is the positive definite diagonal matrix of uniquenesses variances. Loadings and uniquenesses may be constrained across mixture components obtaining the family of 8 possible models, say type C and U, described in Table 1. Following [16], some more flexibility can be achieved by introducing the 4 semi-constrained models, say

type S, in Table 2. In this class of models, the matrix of factor loadings can be written in the form $\mathbf{\Lambda}_g = \mathbf{\Lambda}\mathbf{L}_g$, where $\mathbf{L}_g$ is a positive definite diagonal matrix of factor saliences. They are constrained cases between the first and the last four models in Table 1. The latent factors in each cluster are the same but with different variances recorded by the matrices $\mathbf{L}_g^2$. This is a particular form of factorial invariance firstly introduced by Cattell [5] and then developed by several authors, e.g. in the context of three-way analysis (see [7] and references herein).

Table 1. 8 parsimonious models

Type	Model ID	$\mathbf{\Lambda}_g$	$\mathbf{\Psi}_g$	Isotropic	$\mathbf{\Sigma}_g$
C	CCC	C	C	C	$\mathbf{\Lambda}\mathbf{\Lambda}' + \psi\mathbf{I}_J$
	CCU	C	C	U	$\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$
	CUC	C	U	C	$\mathbf{\Lambda}\mathbf{\Lambda}' + \psi_g\mathbf{I}_J$
	CUU	C	U	U	$\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}_g$
U	UCC	U	C	C	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \psi\mathbf{I}_J$
	UCU	U	C	U	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \mathbf{\Psi}$
	UUC	U	U	C	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \psi_g\mathbf{I}_J$
	UUU	U	U	U	$\mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \mathbf{\Psi}_g$

Table 2. 4 semi-constrained models

Type	Model ID	$\mathbf{\Lambda}_g$	$\mathbf{\Psi}_g$	Isotropic	$\mathbf{\Sigma}_g$
S	SCC	S	C	C	$\mathbf{\Lambda}\mathbf{L}_g^2\mathbf{\Lambda}' + \psi\mathbf{I}_J$
	SCU	S	C	U	$\mathbf{\Lambda}\mathbf{L}_g^2\mathbf{\Lambda}' + \mathbf{\Psi}$
	SUC	S	U	C	$\mathbf{\Lambda}\mathbf{L}_g^2\mathbf{\Lambda}' + \psi_g\mathbf{I}_J$
	SUU	S	U	U	$\mathbf{\Lambda}\mathbf{L}_g^2\mathbf{\Lambda}' + \mathbf{\Psi}_g$

The semi-constrained models have two very interesting properties: (1) They can saturate the model. When $G = 2$ the semi-constrained model can be imposed without loss of fit, i.e. does not impose any restriction. When $G > 2$ it can be imposed without loss of fit if the number of factors can be greater than the number of variables. (2) Factors are unique. Under mild conditions, Kruskal [8] has shown that it is not possible to rotate the factors as in the classical factor analysis model.

4 Total Dimensionality Reduction

There are situations where the clusters lie on a subspace of reduced dimension, i.e. a subspace with a dimension lower than the number of variables. In other

terms, there are noise dimensions in the data that do not contain information about the clustering structure and could even mask it. In order to identify such noise dimensions, we follow [15] and in the general model of Sect. 2 we assume that the J latent variables $\mathbf{y}$ are linear combinations of J factors $\mathbf{f}$, that is $\mathbf{y} = \mathbf{A}\mathbf{f}$ with $\mathbf{A}$ square and non-singular. Such factors can be split into two independent subsets. In the first one, there are $Q(\leq J)$ "discriminative" factors having some clustering information, while in the second set there are $\bar{Q} = J - Q$ factors corresponding to the "noise" dimensions. Technically, the Q discriminative factors $\mathbf{f}^Q$ are distributed as a finite mixture of Gaussians

$$h(\mathbf{f}^Q; \vartheta_{\mathbf{f}}^Q) = \sum_{g=1}^{G} \phi_Q(\mathbf{f}^Q; \boldsymbol{\eta}_g, \boldsymbol{\Omega}_g) p_g,$$

while $\mathbf{f}^{\bar{Q}}$ noise factors are independent of $\mathbf{f}^Q$ and their distribution does not vary from one class to another. In particular, they have a multivariate normal distribution with mean $\boldsymbol{\eta}_0$ and variance $\mathbf{I}_{\bar{Q}}$. It follows that $\mathbf{y}$ is distributed like the finite mixture (1) with

$$\boldsymbol{\mu}_g = E(\mathbf{y}|g) = \mathbf{A}E(\mathbf{f}|g) = \mathbf{A}\begin{bmatrix}\boldsymbol{\eta}_g \\ \boldsymbol{\eta}_0\end{bmatrix} = \mathbf{A}_Q\boldsymbol{\eta}_g + \mathbf{A}_{\bar{Q}}\boldsymbol{\eta}_0 \tag{5}$$

where $\mathbf{A} = [\mathbf{A}_Q, \mathbf{A}_{\bar{Q}}]$, and

$$\boldsymbol{\Sigma}_g = V(\mathbf{y}|g) = \mathbf{A}V(\mathbf{f}|g)\mathbf{A}' = \mathbf{A}\begin{bmatrix}\boldsymbol{\Omega}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\bar{Q}}\end{bmatrix}\mathbf{A}' = \mathbf{A}_Q\boldsymbol{\Omega}_g\mathbf{A}_Q' + \mathbf{A}_{\bar{Q}}\mathbf{A}_{\bar{Q}}'. \tag{6}$$

This model helps in reducing the number of parameters and in finding the least important variables for the classification. Indeed, the total variance of $\mathbf{y}$ can be decomposed into the part explained by the informative and the non-informative factors as

$$V(\mathbf{y}) = V(\mathbf{A}\mathbf{f}) = V(\mathbf{A}_Q\mathbf{f}^Q + \mathbf{A}_{\bar{Q}}\mathbf{f}^{\bar{Q}}) = \mathbf{A}_Q V(\mathbf{f}^Q)\mathbf{A}_Q' + \mathbf{A}_{\bar{Q}}V(\mathbf{f}^{\bar{Q}})\mathbf{A}_{\bar{Q}}'. \tag{7}$$

Variables scarcely related with the discriminative factors, i.e. the ones having only a small part of the variance explained by $\mathbf{f}^Q$, would be the ones less informative for the classification.

5 Between Dimensionality Reduction of Three-Way Data

Let $\mathbf{x} = [x_{11}, \ldots, x_{J1}, \ldots, x_{1K}, \ldots, x_{JK}]'$ be a random vector of J ordinal variables observed at K different occasions. To reduce the number of parameters, we assume a homoscedastic model where the component-specific covariance matrix is modelled as [3]

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_O \otimes \boldsymbol{\Sigma}_V, \tag{8}$$

where $\otimes$ is the Kronecker product of matrices and $\boldsymbol{\Sigma}_O$ and $\boldsymbol{\Sigma}_V$ represent the within covariance matrices of occasions and variables, respectively. The dimensionality reduction is performed on the component-specific mean vectors following a Tucker2 model [19][17]. In details, there are a $J \times Q$ $(\leq J)$ matrix $\mathbf{B}$, a

$K \times R$ ($\leq K$) matrix $\mathbf{C}$ and a $QR \times 1$ vector $\boldsymbol{\xi}_g$ such that, for $g = 1, 2, \ldots, G$, we can write

$$\boldsymbol{\mu}_g = \boldsymbol{\mu} + (\mathbf{C} \otimes \mathbf{B})\boldsymbol{\xi}_g, \tag{9}$$

where $\boldsymbol{\xi}_g$ is such that $\sum_g \boldsymbol{\xi}_g p_g = 0$. It collects the scores of the g-th group on the Q latent variables under the R latent occasions. $\mathbf{B}$ is the loadings matrix that connects the J variables with the Q latent variables and $\mathbf{C}$ is the loadings matrix that connects the K occasions with the R latent occasions. This trilinear model allows us to project the within-group means, lying into a JK dimensional space, onto a reduced subspace of dimension QR.

It is interesting to note that when $K = 1$ this model is equivalent to the homoscedastic version of the model presented in Sect. 4. We can show the equivalence by noting that a set of parameters for the model in Sect. 4 can be written in the form of the parameters of the model in Sect. 5 and vice versa. First, we note that (8) can be obtained from (6) simply by setting

$$\boldsymbol{\Sigma}_g = \mathbf{A} \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\bar{Q}} \end{bmatrix} \mathbf{A}' = \boldsymbol{\Sigma}_V.$$

Second, by rewriting (5) as

$$\begin{aligned} \boldsymbol{\mu}_g &= \mathbf{A} \begin{bmatrix} \boldsymbol{\eta}_g \\ \boldsymbol{\eta}_0 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \sum_g \boldsymbol{\eta}_g p_g \\ \boldsymbol{\eta}_0 \end{bmatrix} + \mathbf{A} \begin{bmatrix} \boldsymbol{\eta}_g - \sum_g \boldsymbol{\eta}_g p_g \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{A} \begin{bmatrix} \sum_g \boldsymbol{\eta}_g p_g \\ \boldsymbol{\eta}_0 \end{bmatrix} + \mathbf{A}_Q (\boldsymbol{\eta}_g - \sum_g \boldsymbol{\eta}_g p_g), \end{aligned}$$

equation (9) is obtained by setting

$$\boldsymbol{\mu} = \mathbf{A} \begin{bmatrix} \sum_g \boldsymbol{\eta}_g p_g \\ \boldsymbol{\eta}_0 \end{bmatrix}, \mathbf{B} = \mathbf{A}_Q, \boldsymbol{\xi}_g = \boldsymbol{\eta}_g - \sum_g \boldsymbol{\eta}_g p_g.$$

On the other side, let $\bar{\mathbf{B}}$ an arbitrary matrix such that $\mathbf{B}_\mathrm{c} = [\bar{\mathbf{B}}, \mathbf{B}]$ is square non singular and $\mathbf{T}$ the lower triangular matrix of the Cholesky decomposition $\mathbf{T}\mathbf{T}' = \mathbf{B}_\mathrm{c}^{-1} \boldsymbol{\Sigma}_V (\mathbf{B}_\mathrm{c}')^{-1}$. We obtain (6) from (8) by writing

$$\begin{aligned} \boldsymbol{\Sigma}_V &= \mathbf{B}_\mathrm{c} \mathbf{T}\mathbf{T}' \mathbf{B}_\mathrm{c}' = \mathbf{B}_\mathrm{c} \begin{bmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \begin{bmatrix} \mathbf{T}_1' & \mathbf{T}_2' \\ \mathbf{0} & \mathbf{T}_3' \end{bmatrix} \mathbf{B}_\mathrm{c}' \\ &= \left[\bar{\mathbf{B}}\mathbf{T}_1 + \mathbf{B}\mathbf{T}_2, \mathbf{B}\mathbf{T}_3\right] \begin{bmatrix} (\bar{\mathbf{B}}\mathbf{T}_1 + \mathbf{B}\mathbf{T}_2)' \\ \mathbf{T}_3' \mathbf{B}' \end{bmatrix} \\ &= (\bar{\mathbf{B}}\mathbf{T}_1 + \mathbf{B}\mathbf{T}_2)(\bar{\mathbf{B}}\mathbf{T}_1 + \mathbf{B}\mathbf{T}_2)' + \mathbf{B}\mathbf{T}_3 \mathbf{T}_3' \mathbf{B}' \end{aligned}$$

and setting

$$\mathbf{A}_Q = \mathbf{B}, \mathbf{A}_{\bar{Q}} = \bar{\mathbf{B}}\mathbf{T}_1 + \mathbf{B}\mathbf{T}_2, \boldsymbol{\Omega} = \mathbf{T}_3 \mathbf{T}_3'.$$

For the conditional means, (5) can be obtained from (9) by writing $\mathbf{A}^{-1}\boldsymbol{\mu} = [\zeta'_Q, \zeta'_{\bar{Q}}]'$ and

$$\boldsymbol{\mu}_g = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi}_g = \mathbf{A}\mathbf{A}^{-1}\boldsymbol{\mu} + \mathbf{A}_Q\boldsymbol{\xi}_g = \mathbf{A}_Q(\zeta_Q + \boldsymbol{\xi}_g) + \mathbf{A}_{\bar{Q}}\zeta_{\bar{Q}}$$

and by setting

$$\boldsymbol{\eta}_g = \zeta_Q + \boldsymbol{\xi}_g, \boldsymbol{\eta}_0 = \zeta_{\bar{Q}}.$$

6 Discussion

In this paper, we discussed possible approaches to cluster ordinal data. We have seen that the problem is rather complex, both from a computational and a modeling point of view. The main idea is to interpret the ordinal variables as a discretization of some latent continuous variables. This makes it possible to draw ideas from the vast existing literature on the continuous case. It is important to say that this approach can be extended to the case of mixed data, i.e. some variables ordinal some others continuous, in a quite natural way [14].

Computationally, the complexity can be reduced by adopting a composite likelihood approach. Ranalli and Rocci [13,15] considered a surrogate function build on all the bivariate marginals. The statistical efficiency could be improved, at the cost of an increment in the computational complexity, by considering marginals of higher order. Possible extensions to the mixed case have been proposed by [14,16].

Three parsimonious reparameterizations have been discussed and all are based on the dimensionality reduction of the data. In the first, the reduction is performed within the clusters, in the second it is performed in the total space while in the third between the clusters. The choice between the different approaches depends on the objective of the analysis. For example, the within model should be adopted if we suspect different factors within the clusters. The between approach is appropriate to select or improve the variables. It identifies dimensions containing the information about the clustering structure if it is defined only by the component means. Whenever we think that the clusters differs also in shape, i.e. covariance structure, then the reduction in the total space should be preferred. Of course, the use of one approach does not exclude the others, i.e., the different approaches could be joined into a unique model.

As a final remark, we say that it is important to check the identifiability of the model taking into account, on one side the rotational freedom of the factors, and on the other the reduced number of estimable parameters due to the use of the composite likelihood. For more details see [13,15].

References

1. Bartholomew, D.J., Knott, M., Moustaki, I.: Latent Variable Models and Factor Analysis: A Unified Approach, 3rd edn. Wiley Series in Probability and Statistics, Wiley (2011)
2. Bock, D., Moustaki, I.: Item response theory in a general framework. In: Rao, C., Sinharay, S. (Eds.) Handbook of Statistics on Psychonometrics, Chap. 26. Elsevier (2007)
3. Browne, M.W.: The decomposition of multitrait-multimethod matrices. British J. Math. Stat. Psychol. **37**(1), 1–21 (1984)
4. Cagnone, S., Viroli, C.: A factor mixture analysis model for multivariate binary data. Stat. Model. **12**, 257–277 (2012)
5. Cattell, R.B.: "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. Psychometrika **9**(4), 267–283 (1944)
6. Ghahramani, Z., Hinton, G.E.: The EM algorithm for mixtures of factor analyzers (1996). Available at https://api.semanticscholar.org/CorpusID:18270595
7. Giordani, P., Rocci, R., Bove, G.: Factor uniqueness of the structural parafac model. Psychometrika **85**(3), 555–574 (2020)
8. Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl. **18**(2), 95–138 (1977)
9. Lindsay, B.: Composite likelihood methods. Contemp. Math. **80**, 221–239 (1988)
10. McLachlan, G., Peel, D., Bean, R.: Modelling high-dimensional data by mixtures of factor analyzers. Comput. Stat. Data Anal. **41**(3–4), 379–388 (2003)
11. McLachlan, G.J., Rathnayake, S.I.: Mixture models for standard p-dimensional euclidean data. In: Hennig, C., Meila, M., Rocci, R., Murtagh, F. (Eds.) Handbook of Cluster Analysis, Chap. 8. Chapman and Hall/CRC (2015)
12. Muthén, B.: A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika **49**(1), 115–132 (1984)
13. Ranalli, M., Rocci, R.: Mixture models for ordinal data: a pairwise likelihood approach. Stat. Comput. **26**(1), 529–547 (2016)
14. Ranalli, M., Rocci, R.: Mixture models for mixed-type data through a composite likelihood approach. Comput. Stat. Data Anal. 110(C), 87–102 (2017a). https://doi.org/10.1016/j.csda.2016.12.01
15. Ranalli, M., Rocci, R.: A model-based approach to simultaneous clustering and dimensional reduction of ordinal data. Psychometrika **82**(4), 1007–1034 (2017). https://doi.org/10.1007/s11336-017-9578-5
16. Ranalli, M., Rocci, R.: Composite likelihood methods for parsimonious model-based clustering of mixed-type data. Adv. Data Anal. Classif. (2023). https://doi.org/10.1007/s11634-023-00539-5
17. Rocci, R., Vichi, M.: Three-mode component analysis with crisp or fuzzy partition of units. Psychometrika **70**(4), 715–736 (2005)
18. Takane, Y., de Leeuw, J.: On the relationship between item response theory and factor analysis of discretized variables. Psychometrika **52**(3), 393–408 (1987)
19. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
20. Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Statist. Sinica **21**(1), 1–41 (2011)

Check for updates

Robustness in LiNGAM

Jakob Raymaekers(✉)

Department of Mathematics, University of Antwerp, Antwerp, Belgium
jakob.raymaekers@uantwerpen.be

Abstract. We review the LiNGAM structural causal model, its identifiability properties and the three most popular algorithms for fitting the model on real data. We illustrate the sensitivity of these algorithms to a single outlier in the data. A robust counterpart for each of the three algorithms is presented, and it is shown that one of them, TSLiNGAM, is particularly promising.

Keywords: Structural causal models · Causal structure learning · Robustness

1 Causal Discovery and LiNGAM

Assume we observe a p-dimensional random vector $\boldsymbol{X}$ and suppose that the distribution of $\boldsymbol{X}$, denoted $\mathbb{P}_{\boldsymbol{X}}$, is entailed by an underlying structural causal model (SCM) $\mathcal{C}(\boldsymbol{S}, \mathbb{P}_{\boldsymbol{N}})$ [13]. We can characterize the SCM by a pair $(\boldsymbol{S}, \mathbb{P}_{\boldsymbol{N}})$ where $\mathbf{S}$ is a set of structural assignments

$$X_i = f_i(\mathbf{PA}_i, N_i) \qquad i = 1, \ldots, p.$$

For each i, $\mathbf{PA}_i \subset \{X_1, \ldots, X_p\} \backslash \{X_i\}$ denote the parents of X_i, and N_i denotes the noise variable connected to variable i. $\mathbb{P}_{\boldsymbol{N}}$ denotes the joint distribution over the noise variables $\boldsymbol{N} = (N_1, \cdots, N_p)$ which are assumed to be jointly independent. We further assume that the ordering of structural assignments can be represented by a directed acyclic graph. This also implies that there is always a *causal ordering* of the variables, i.e. a permutation $\pi(1), \ldots, \pi(p)$ of the variables so that $\mathbf{PA}_i \subset \{X_k | \pi(k) < \pi(i)\}$. Finally, we assume that for each i, the corresponding noise variable N_i is independent from the parents $\mathbf{PA}_i$ of X_i.

In causal discovery, we aim to recover the underlying causal graph based on observed samples from the distribution $\mathbb{P}_{\boldsymbol{X}}$. In full generality, this task often runs into identifiability issues. More precisely, we face the problem that different SCMs induce the same joint probability distribution of $\boldsymbol{X}$, making it impossible to infer the SCM that generated the data from the observed samples alone. As a simple bivariate example, suppose $X_1 \sim \mathcal{N}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,1)$ are independent, and $X_2 = 2X_1 + \varepsilon_2$. Then $\boldsymbol{X} \sim \mathcal{N}\left(\mathbf{0}_2, \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}\right)$, and the SCM $X_1 = \frac{2}{\sqrt{5}} X_2 + \frac{1}{\sqrt{5}} \varepsilon_1$ with independent $X_2 \sim \mathcal{N}(0,1)$ and $\varepsilon_1 \sim \mathcal{N}(0,1)$ generates

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 388–395, 2024.
https://doi.org/10.1007/978-3-031-65993-5_48

the same joint distribution. More generally, if the functions f_i are linear and the noise variables are Gaussian, then the underlying SCM is unidentifiable.

One approach to obtaining the identifiability of the underlying SCM, is to place further restrictions on the nature of the SCM. The most popular way of doing this is by restricting the distribution $\mathbb{P}_{\boldsymbol{N}}$ of the noise variables, and/or restricting the class of functions $\mathcal{F}$ to which the f_i must belong. Possibly the earliest result in this direction was the introduction of the *Linear Non-Gaussian Additive Model* (LiNGAM) [16]. As the name suggests, the LiNGAM family contains all SCMs for which the function class $\mathcal{F}$ is restricted to linear functions, and the noise variables $N_1, \ldots, N_p$ are required to be continuous, non-Gaussian random variables. For simplicity of notation, we assume that every variable has zero mean. The LiNGAM structure then implies that each variable i can be written as

$$X_i = \sum_{j:\pi(j)<\pi(i)} \beta_{ij} X_j + N_i.$$

This implies in turn that we can write

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{B}$ is a matrix of coefficients that can be permuted (by simultaneous row and column permutations) to a strictly lower triangular matrix. The actual permutation which obtains such a strict lower-triangular $\boldsymbol{B}$ is determined by the true causal ordering. It turns out that SCMs from a LiNGAM family are indeed identifiable from the joint distribution $\mathbb{P}_{\boldsymbol{X}}$.

2 Algorithms

We now discuss some of the most popular algorithms for estimating a structural causal model. It should be noted that the hardest part of estimating a SCM is finding the correct causal order. Once such an order has been found, the estimation of the coefficients can be achieved with appropriate regression techniques [1,19]. The relevant regression technique for estimating the coefficients is dictated by the assumptions made on the form of the structural assignments. Additionally, variable selection is usually necessary to discard any non-existing connections between lower and higher-ordered variables. We focus here on identifying the causal order.

2.1 ICA-LiNGAM

The first algorithm for estimating the causal order for a SCM from the LiNGAM family was introduced by [16]. Note that we can rewrite Eq. 1 as $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{N}$ where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$. This equation, together with the assumption that $\boldsymbol{N}$ has non-Gaussian independent components (technically, a single Gaussian noise variable is allowed), is the well-known *independent component analysis* (ICA) model introduced by [2]. The matrix $\boldsymbol{A}$ (and thus also $\boldsymbol{B}$) is identifiable upon scaling

and permutation of the columns. The algorithm for estimating the LiNGAM SCM introduced by [16] builds on this idea and is therefore often referred to as *ICA-LiNGAM*. It starts out by applying ICA to the standardized data. This yields an estimate of $\boldsymbol{I} - \boldsymbol{B}$, of which the rows are then permuted so that the right noise vector corresponds with the right variable. This produces an estimate of $\boldsymbol{B}$, which in the final step is then permuted into a strictly lower triangular matrix using simultaneous row- and column permutations.

2.2 DirectLiNGAM

DirectLiNGAM [17] is an alternative algorithm for estimating a LiNGAM SCM. Instead of relying on ICA, it is a stepwise algorithm that makes use of simple linear regressions and independence tests. In each step, it identifies an *exogenous variable*, i.e. a variable without parents. This variable is then removed from the active set and all remaining variables are replaced by their residuals after regressing on the identified exogenous variable. When no variables remain in the active set, the order in which the exogenous variables were identified is the estimated causal order. DirectLiNGAM builds crucially on the following two Lemmas

Lemma 1. *(***Lemma 1 of** *[17]) Assume that* $\boldsymbol{X}$ *strictly follows the LiNGAM model. Let* $R_i^{(j)} = X_i - \frac{\mathrm{Cov}[X_i, X_j]}{\mathrm{Var}[X_j]} X_j$ *be the residual when* X_i *is regressed on* X_j*. Then a variable* X_j *is exogenous iff* X_j *is independent of its residuals* $R_i^{(j)}$ *for all* $i \neq j$*.*

Lemma 2. *(***Lemma 2 of** *[17]) Assume that* $\boldsymbol{X}$ *strictly follows the LiNGAM model and suppose* X_j *is an exogenous variable. Denote by* $\boldsymbol{R}^{(j)}$ *the* $p-1$*-dimensional vector collecting the residuals* $R_i^{(j)}$ *when all* X_i *of* $\boldsymbol{X}$ *are regressed on* X_j *(*$i \neq j$*). Then a LiNGAM model holds for the residual vector* $\boldsymbol{R}^{(j)}$*.*

Lemma 1 ensures that we can identify an exogenous variable by verifying the independence of the variable and the residuals of simple linear regressions. Lemma 2 guarantees that when an exogenous variable is identified, we can indeed remove it from the active set, replace the remaining variables by their residuals from regressing on the removed variable, and continue with the iterative process. The advantage of DirectLiNGAM over ICA-LiNGAM is that given enough data, it is guaranteed to converge to the right solution in a finite number of steps. A potential drawback is that the computation time does not scale well with the number of variables.

2.3 Pairwise LiNGAM

The final algorithm we discuss here is the *Pairwise LiNGAM* algorithm [6]. As the name suggests, Pairwise LiNGAM relies on a likelihood-ratio-based measure of the causal direction between a pair of variables. Consider a bivariate random

vector $\boldsymbol{X} = (X_1, X_2)$ generated by a LiNGAM model and assume that the variables have mean zero and variance one. Based on samples from the distribution of $\boldsymbol{X}$ we want to decide which of the following two models generated the data:

$$X_1 \to X_2 : X_2 = \rho X_1 + N_2$$
$$X_2 \to X_1 : X_1 = \rho X_2 + N_1,$$

where ρ denotes the correlation coefficient. Now denote the likelihood of the LiNGAM with $X_1 \to X_2$ as $L(X_1 \to X_2)$, and similarly as $L(X_2 \to X_1)$ for the model $X_2 \to X_1$. For sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ where $\boldsymbol{x}_i = (x_{i,1}, x_{i,2})$ for $i = 1, \ldots, n$ we can define the log-likelihood-ratio as [6]

$$\begin{aligned} R_{12} =& \frac{1}{n} \log L(X_1 \to X_2) - \frac{1}{n} \log L(X_2 \to X_1) \\ =& \frac{1}{n} \sum_{i=1}^{n} G_{X_1}(x_{i,1}) + G_{N_1}\left(\frac{x_{i,2} - \rho x_{i,1}}{\sqrt{1-\rho^2}}\right) - G_{X_2}(x_{i,2}) + G_{N_2}\left(\frac{x_{i,1} - \rho x_{i,2}}{\sqrt{1-\rho^2}}\right), \end{aligned}$$

where G_{X_1}, G_{X_2} denote the log-pdf of X_1 and X_2 respectively, and G_{N_1}, G_{N_2} denote the standardized log-pdfs of the residuals when regression X_1 on X_2 and X_2 on X_1 respectively. If R is positive, we conclude $X_1 \to X_2$, otherwise we conclude $X_2 \to X_1$. In practice, the different log-pdfs $G_{X_1}, G_{X_2}, G_{N_1}, G_{N_2}$ and the correlation ρ need to be estimated. The latter can be done by maximum likelihood or through least-squares regression. For the log-pdfs, various parametric and non-parametric approaches could be considered. For small sample sizes, this may be difficult however, which is why [6] advocate the use of a first-order approximation of the likelihood ratio above assuming a logistic density for G which yields the much simpler $\tilde{R}_{12} = \rho \sum_{i=1}^{n} x_{i,1} \tanh(x_{i,2}) - x_{i,2} \tanh(x_{i,1})$.

The pairwise log-likelihood ratio can be leveraged in an algorithm for identifying general multivariate LiNGAM models. The first way of doing this is through a stepwise algorithm in the same spirit as DirectLiNGAM. At each iteration a root of the graph is found by finding the variable i for which all $\tilde{R}_{ij}$ $j \neq i$ are positive. A second way of incorporating the pairwise log-likelihood ratios is by first estimating the *skeleton* of the DAG (i.e., the set of undirected edges), and in a second step identifying the direction of each edge. This works particularly well if the DAG underlying the SCM is sparse. For estimating the initial skeleton, [6] propose to use the precision matrix with a bootstrap test for identifying the true zeroes.

3 Robustness in Causal Discovery

Robustness has received little to no attention in the causal discovery literature so far. There are several potential reasons for this. Causal discovery is an area of research that has blossomed only recently and robustness considerations are typically put forward after the classical non-robust cases have been treated to a satisfactory level. Additionally, there are intuitive reasons why robustness may be particularly challenging in the context of causal discovery and LiNGAM models.

It should be clear from the discussion in Sects. 1 and 2 that the LiNGAM model is fundamentally connected to the concept of (joint) independence as the observed data is a linear transformation of jointly independent variables. Recently, the research around measures of (joint) independence has gained increasing attention [4,14,18] and it is very much an active area of research in progress. Robustness has not really been treated in the literature on independence measures, with exception of the recent study and proposal in [8].

To illustrate the lack of robustness of the most popular LiNGAM algorithms, we do the following experiment, inspired by [7]. We generate samples of size $n = 1000$ of a bivariate random vector (X, Y) where $X \sim t_5$ and $Y = X + \varepsilon$ where $\varepsilon \sim t_5$ and $X \perp\!\!\!\perp \varepsilon$. In each sample, we replace one observation by the pair (z_1, z_2), which we vary over a grid, and we verify whether the different algorithms correctly infer the causal direction $X \to Y$. We repeat the experiment 100 times for each value of (z_1, z_2) and compute the fraction of correctly identified causal directions. For ICA-LiNGAM, we use the implementation in the `R`-package `pcalg` [9]. For pairwise LiNGAM, we use the `R`-package `causalXtreme` [3].

The results of this experiment are presented in Fig. 1. It is clear that for all three of the algorithms discussed so far, a single outlier can flip the estimated direction of causality. Additionally, it seems that ICA-LiNGAM behaves somewhat differently from the other two in its sensitivity to outliers. DirectLiNGAM and pairwise LiNGAM tend to wrongly estimate the causal direction for the same values of the outlier (z_1, z_2). This can be explained by the fact that these two algorithms both work in a stepwise fashion, unlike ICA-LiNGAM. In any case, the fact that a single contaminated point can yield a result that directly opposes the truth is worrisome.

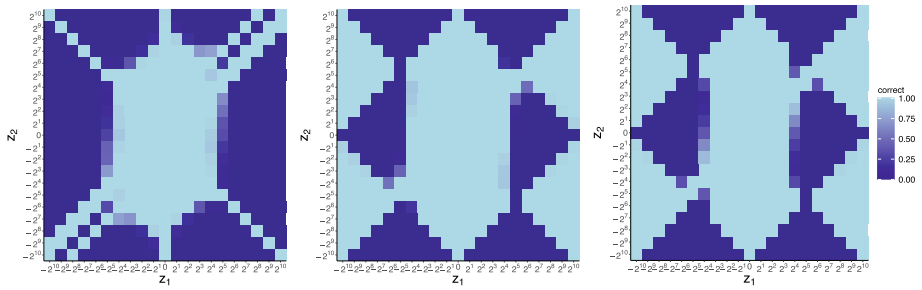


Fig. 1. Fraction of correctly estimated causal directions when one outlier is placed at (z_1, z_2) for ICA-LiNGAM (left), DirectLiNGAM (center) and pairwise LiNGAM (right).

4 Robustified Algorithms

In this section, we consider three versions of the above algorithms which are tailored towards heavier-tailed distributions and hopefully guard somewhat against

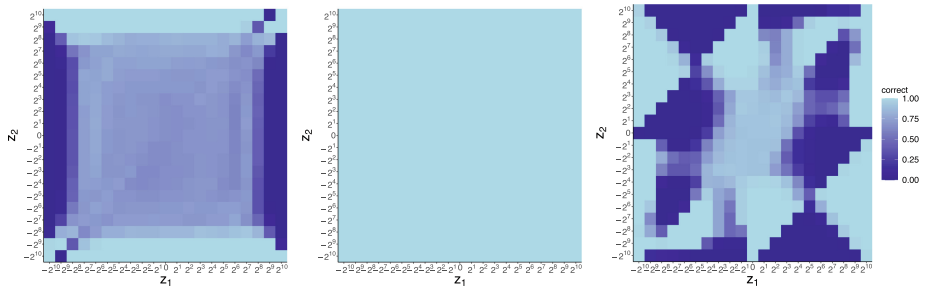


Fig. 2. Fraction of correctly estimated causal directions when one outlier is placed at (z_1, z_2) for robust ICA-LiNGAM (left), TSLiNGAM (center) and robust pairwise LiNGAM (right).

outliers. To robustify ICA-LiNGAM, we consider replacing the fastICA step by a robust independent component analysis based on the joint diagonalization of Tyler's shape matrix and Duembgen's shape matrix [10]. For this, we use the `R`-package `ICS` [11]. For DirectLiNGAM, we consider the recent proposal TSLiNGAM [7] which leverages the robustness and efficiency of the Theil-Sen slope estimator [15,20] in the original DirectLiNGAM algorithm. Finally, for pairwise LiNGAM, we consider the version with the robust approximation of the likelihood ratio given by $\tilde{R}_{12} = \rho \sum_{i=1}^{n} g(x_{i,1})x_{i,2} - x_{i,1}g(x_{i,2})$ with $g(t) = \log(\cosh(\max\{0, t\}))$ [6].

Repeating the experiment of Sect, 3 with the robustified algorithms yields Fig. 2. The robust ICA-LiNGAM is more robust than the original ICA-LiNGAM. However, it seems to be somewhat unstable as even on uncontaminated data, it estimates the wrong causal direction around 20 to 30% of the time. This approach can possibly be improved upon by a detailed comparative study of the different scatter matrices used for diagonalization. The TSLiNGAM algorithm is not influenced by the single outlier, and always estimates the correct causal direction in this setup. For the robust pairwise LiNGAM, the pattern has changed somewhat compared with Fig. 1 but the overall robustness is very similar, in that a single outlier can easily flip the estimated direction of causality.

5 Conclusion

Robustness in causal discovery is a challenging subject. For the LiNGAM family of structural causal models, the most popular algorithms are extremely sensitive to outlying values as a single extreme observation can flip the estimated direction of causality. The recent TSLiNGAM [7] is shown to have promising robustness properties. Further research for robust causal discovery in LiNGAM models is needed to investigate how this robustness carries over to higher-dimensional models, how ICA-LiNGAM and pairwise LiNGAM can be robustified more efficiently, and how these robustifications may impact the efficiency of the methods on small samples. Of course, the LiNGAM assumption represents only one family

of models in which the underlying SCM can be identified. Robustness in nonlinear SCMs, such as the nonlinear additive noise models [5,12] provide a wealth of future challenges.

Acknowledgements. The author was supported by the European Union's Horizon 2022 research and innovation program under the Marie Sklodowska Curie grant agreement No 101103017.

References

1. Bühlmann, P., Peters, J., Ernest, J.: CAM: causal additive models, high-dimensional order search and penalized regression. Ann. Statist. **42**(6), 2526–2556 (2014). https://doi.org/10.1214/14-AOS1260
2. Comon, P.: Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
3. Gnecco, N., Meinshausen, N., Peters, J., Engelke, S.: causalXtreme: causal discovery in heavy-tailed models. R package version 0.0.0.9000 (2024)
4. Gretton, A., Bousquet, O., Smola, A., Schölkopf, B.: Measuring statistical dependence with hilbert-schmidt norms. In: International Conference on Algorithmic Learning Theory, pp. 63–77. Springer (2005)
5. Hoyer, P., Janzing, D., Mooij, J.M., Peters, J., Schölkopf, B.: Nonlinear causal discovery with additive noise models. In: Advances in Neural Information Processing Systems 21 (NIPS), pp. 689–696 (2009)
6. Hyvärinen, A., Smith, S.M.: Pairwise likelihood ratios for estimation of non-gaussian structural equation models. J. Mach. Learn. Res. **14**(1), 111–152 (2013)
7. Leyder, S., Raymaekers, J., Verdonck, T.: Tslingam: Directlingam under heavy tails. arXiv:2308.05422 (2023)
8. Leyder, S., Raymaekers, J., Rousseeuw, P.J.: Is distance correlation robust? arXiv:2403.03722 (2024)
9. Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., Bühlmann, P.: Causal Inference Using Graphical Models with the R Package pcalg. J. Stat. Softw. **47**(11), 1–26 (2012). https://doi.org/10.18637/jss.v047.i11
10. Nordhausen, K., Oja, H., Ollila, E.: Robust independent component analysis based on two scatter matrices. Austrian J. Stat. **37**(1), 91–100 (2008)
11. Nordhausen, K., Oja, H., Tyler, D.E.: Tools for exploring multivariate data: the package ICS. J. Stat. Softw. **28**(6), 1–31 (2008)
12. Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B.: Causal discovery with continuous additive noise models. J. Mach. Learn. Res. **15**(58):2009–2053 (2014). http://jmlr.org/papers/v15/peters14a.html
13. Peters, J., Janzing, D., Schölkopf, B.: Elements of causal inference: foundations and learning algorithms. The MIT Press, Cambridge, MA (2017)
14. Pfister, N., Bühlmann, P., Schölkopf, B., Peters, J.: Kernel-based tests for joint independence. J. R. Stat. Soc. Ser. B Stat. Method. **80**(1), 5–31 (2018)
15. Sen, P.K.: Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. **63**(324), 1379–1389 (1968)
16. Shimizu, S., Hoyer, P.O., Hyvarinen, A., Kerminen, A.: A linear non-gaussian acyclic model for causal discovery. J. Mach. Learn. Res. **7**, 2003–2030 (2006)
17. Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P.O., Bollen, K.: DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model. J. Mach. Learn. Res. **12**, 1225–1248 (2011)

18. Székely, G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. Ann. Statist. **35**(6), 2769–2794 (2007). https://doi.org/10.1214/009053607000000505
19. Teyssier, M., Koller, D.: Ordering-based search: a simple and effective algorithm for learning bayesian networks. In: Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence, UAI'05, pp. 584-590. AUAI Press, Arlington, Virginia, USA (2005)
20. Theil, H. A rank-invariant method of linear and polynomial regression analysis, III. In: Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen (1950)

Comparing Bivariate Random Vectors by Means of Statistical Preference

Julián Ros, Raúl Pérez-Fernández(✉), and Ignacio Montes

Department of Statistics and O.R., University of Oviedo, C/Federico García Lorca 18, 33007 Oviedo, Spain
{uo277560,perezfernandez,imontes}@uniovi.es

Abstract. Stochastic orders are probabilistic tools used for comparing random quantities. Statistical preference is a stochastic ordering with two main features: firstly, the comparison is based on the joint distribution of the random variables, thus considering potential dependence between them. Secondly, it is accompanied by a winning probability, which measures the strength of the preference. This contribution proposes several ways to extend statistical preference for comparing (bivariate) random vectors. These proposals follow three different approaches: (i) aggregating the winning probabilities of each component, (ii) computing the winning probabilities between the aggregated random vectors, or (iii) defining a purely bivariate winning probability. Besides investigating these properties, the relationships between them and the componentwise median are studied.

Keywords: Stochastic order · Statistical preference · Precedence order · Multivariate statistics

1 Introduction

The comparison of random variables naturally appears in fields such as decision making, finance and economics. Commonly, this comparison is made by means of stochastic orders (see e.g. [1,12]). Among the many stochastic orders that can be found in the literature, *stochastic dominance* [7] is the most widespread. It is based on the comparison of the cumulative distribution functions of the random variables to be compared. Although it has been widely applied in the literature, it only uses marginal information, ignoring the possible dependence between the random variables and, quite typically, resulting in incomparability.

An alternative that avoids the aforementioned issues is to use statistical preference [4,5]. This stochastic order is based on a probabilistic relation, called *winning probability*, that measures the strength of the preference of one random variable over the other one. In addition, since the winning probability is defined by using the joint distribution between the random variables, the dependence between them, captured by a copula [13], is taken into account. The phylosophy

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 396–404, 2024.
https://doi.org/10.1007/978-3-031-65993-5_49

behind the winning probability goes back to Mann and Withney [8], and it has also been called *measure of stochastic superiority* [14] or *probability index* [6].

In previous contributions, the connection between stochastic dominance and statistical preference has been analysed both under independence and also under some dependence structures [3,10], giving simple expressions of the winning probability for some particular dependence structures. As well, statistical preference has been extended for the comparison of several alternatives giving a multivariate winning probability of a random variable over the others [11].

In this contribution we move towards a different direction: we explore how statistical preference can be extended for the comparison of bivariate random vectors following three different paths. Firstly, we compute the winning probabilities componentwisely and we aggregate them by means of an appropriate aggregation operator. Secondly, we aggregate the components of each random vector giving rise to the aggregated random variables and we compute the winning probability between them. Thirdly, we consider the joint distribution of both random vectors and use it to define a purely bivariate winning probability. In addition to analysing each of these approaches, we also explore the connections between them and also the connection with the componentwise median.

The rest of the paper is organised as follows: after reviewing the definition and main properties of statistical preference in Sect. 2, in Sect. 3 we propose several extensions of statistical preference for the comparison of bivariate random vectors and in Sect. 4 we investigate the connections between the proposed extensions. Even though this paper presents an initial study of the extension of statistical preference for random vectors, we explain in Sect. 5 some of the interesting problems that can be investigated within this framework.

2 Comparison of Random Variables

We consider a probability space (Ω, σ, P), and we consider a set $\mathcal{A}$ of random variables defined from Ω to $\mathbb{R}$. The random variables will be denoted by $\mathbf{x}$, $\mathbf{y}$.

Among the many stochastic orders that can be found in the literature, we focus in this paper in *statistical preference*. For defining this notion, we first need to introduce the function $Q : \mathcal{A} \times \mathcal{A} \to [0, 1]$ given, for any $\mathbf{x}, \mathbf{y} \in \mathcal{A}$, by:

$$Q(\mathbf{x}, \mathbf{y}) = P(\mathbf{x} > \mathbf{y}) + \frac{1}{2} P(\mathbf{x} = \mathbf{y}) \,.$$

This function is a probabilistic or reciprocal relation, meaning that $Q(\mathbf{x}, \mathbf{y}) + Q(\mathbf{y}, \mathbf{x}) = 1$. The value $Q(\mathbf{x}, \mathbf{y})$ is usually referred to as the *winning probability* of $\mathbf{x}$ over $\mathbf{y}$. Statistical preference is defined by using these winning probabilities.

Definition 1. Given two random variables $\mathbf{x}, \mathbf{y} \in \mathcal{A}$, $\mathbf{x}$ is (strictly) *statistically preferred* to $\mathbf{y}$, denoted as $\mathbf{x} \succ_{\text{SP}} \mathbf{y}$, if $Q(\mathbf{x}, \mathbf{y}) > \frac{1}{2}$.

Proposition 1. Given $\mathbf{x}, \mathbf{y} \in \mathcal{A}$, the following conditions for the definition of $\mathbf{x} \succ_{\text{SP}} \mathbf{y}$ are equivalent:

$$Q(\mathbf{x},\mathbf{y}) > 1/2 \Leftrightarrow Q(\mathbf{x},\mathbf{y}) > Q(\mathbf{y},\mathbf{x}) \\ \Leftrightarrow P(\mathbf{x} > \mathbf{y}) > P(\mathbf{y} > \mathbf{x}) \Leftrightarrow P(\mathbf{x} \geq \mathbf{y}) > P(\mathbf{y} \geq \mathbf{x}).$$

In the literature, when two random variables satisfy the fourth item in the previous lemma $\mathbf{x}$ is said to dominate $\mathbf{y}$ by the *precedence order* [2]. However, even if statistical preference and the precedence order are equivalent, statistical preference gives more information because the winning probability measures the strength of the preference of $\mathbf{x}$ over $\mathbf{y}$, in the sense that the larger the winning probability, the stronger the preference.

In [9], the connection between statistical preference and location parameters was analysed. In particular, the following connection between statistical preference and the median holds. From now on, we denote the set of medians of a random variable $\mathbf{x}$ (which is known to be an interval or a singleton) by $\text{Me}(\mathbf{x})$.

Proposition 2. ([9, Thm. 15]) Given two random variables $\mathbf{x}, \mathbf{y} \in \mathcal{A}$, the following statements hold:

1. $\mathbf{x} \succ_{\text{SP}} \mathbf{y} \Rightarrow \inf \text{Me}(\mathbf{x} - \mathbf{y}) \geq 0$.
2. $\inf \text{Me}(\mathbf{x} - \mathbf{y}) > 0 \Rightarrow \mathbf{x} \succ_{\text{SP}} \mathbf{y}$.
3. If $P(\mathbf{x} = \mathbf{y}) = 0$, then $\mathbf{x} \succ_{\text{SP}} \mathbf{y} \Leftrightarrow \inf \text{Me}(\mathbf{x} - \mathbf{y}) > 0$.

3 Extensions of Statistical Preference

In this section we propose several extensions of statistical preference for the comparison of random vectors. For simplicity, we consider bivariate random vectors, but most of these proposals can be straightforwardly extended to a more general framework. Taking this into account, we consider a set $\mathcal{A}_2$ of two-dimensional random vectors, denoted by $\vec{\mathbf{x}}$ or $\vec{\mathbf{y}}$, defined in the probability space (Ω, σ, P). From now on we will use the following notation: given two two-dimensional vectors $\vec{a}, \vec{b}$, $\vec{a} \geq \vec{b}$ will be used for referring to the fact that $a_1 \geq b_1$ and $a_2 \geq b_2$. As well, we will use $\vec{a} > \vec{b}$ when $a_1 > b_1$ and $a_2 > b_2$.

3.1 Aggregation of Winning Probabilities

Our first proposal is to compute the winning probabilities between each components and aggregate them.

Definition 2. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and an idempotent[1] and strictly increasing function $M : [0,1] \times [0,1] \to [0,1]$, we define the aggregated winning probability of $\vec{\mathbf{x}}$ over $\vec{\mathbf{y}}$ as the function $Q^M : \mathcal{A}_2 \times \mathcal{A}_2 \to [0,1]$ given by:

$$Q^M(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = M\big(Q(\mathbf{x_1}, \mathbf{y_1}), Q(\mathbf{x_2}, \mathbf{y_2})\big)\,.$$

[1] A function $M : [0,1] \times [0,1] \to [0,1]$ is idempotent if $M(x,x) = x$ for any $x \in [0,1]$.

The function Q^M is a probabilistic relation when we impose additional conditions on the function M.

Proposition 3. If M is such that $M(a,b) + M(1-a,1-b) = 1$ for any $a, b \in [0,1]$, then Q^M is a probabilistic relation.

By using the aggregated winning probability, we can extend statistical preference to the bivariate framework.

Definition 3. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and an idempotent and strictly increasing function $M : [0,1] \times [0,1] \to [0,1]$, we say that $\vec{\mathbf{x}}$ is Q^M-statistically preferred to $\vec{\mathbf{y}}$, denoted as $\vec{\mathbf{x}} \succ_{\text{SP}}^{M} \vec{\mathbf{y}}$, if $Q^M(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > \frac{1}{2}$.

One particular case that can be embedded in this framework is by considering the arithmetic mean. If we denote $\text{am}(a,b) = \frac{1}{2}(a+b)$, we get:

$$Q^{\text{am}}(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = \frac{1}{2}\big(Q(\mathbf{x_1}, \mathbf{y_1}) + Q(\mathbf{x_2}, \mathbf{y_2})\big)\,.$$

As the arithmetic mean verifies the condition in Proposition 3, Q^{am} is a probabilistic relation. Thus, $\vec{\mathbf{x}} \succ_{\text{SP}}^{\text{am}} \vec{\mathbf{y}}$ holds if the aggregated winning probability of the components is greater than or equal to $\frac{1}{2}$. As well, we get the following conditions.

Proposition 4. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$, the following conditions for the definition of $\vec{\mathbf{x}} \succ_{\text{SP}}^{\text{am}} \vec{\mathbf{y}}$ are equivalent:

$$\begin{aligned} Q^{\text{am}}(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > 1/2 &\Leftrightarrow Q^{\text{am}}(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > Q^{\text{am}}(\vec{\mathbf{y}}, \vec{\mathbf{x}}) \\ &\Leftrightarrow P(\mathbf{x_1} > \mathbf{y_1}) + P(\mathbf{x_2} > \mathbf{y_2}) > P(\mathbf{y_1} > \mathbf{x_1}) + P(\mathbf{y_2} > \mathbf{x_2}) \\ &\Leftrightarrow P(\mathbf{x_1} \geq \mathbf{y_1}) + P(\mathbf{x_2} \geq \mathbf{y_2}) > P(\mathbf{y_1} \geq \mathbf{x_1}) + P(\mathbf{y_2} \geq \mathbf{x_2}) \end{aligned}$$

3.2 Winning Probability for the Aggregated Components

The first approach computes the winning probability for each component of the random vector and then aggregates them. Instead, we could first aggregate the components of each random vector, giving rise to random variables, and later compute the (univariate) winning probability. Formally:

Definition 4. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and a strictly increasing function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we define the winning probability of the aggregated components of $\vec{\mathbf{x}}$ over those of $\vec{\mathbf{y}}$ as the function $Q^g : \mathcal{A}_2 \times \mathcal{A}_2 \to [0,1]$ given by:

$$Q^g(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = Q\big(g(\vec{\mathbf{x}}), g(\vec{\mathbf{y}})\big) = P\big(g(\vec{\mathbf{x}}) > g(\vec{\mathbf{y}})\big) + \frac{1}{2}P\big(g(\vec{\mathbf{x}}) = g(\vec{\mathbf{y}})\big)\,.$$

Note that strict increasingness of the function g implies measurability, which eventually guarantees that $g(\vec{\mathbf{x}})$ and $g(\vec{\mathbf{y}})$ are random variables. Additionally, Q^g is a probabilistic relation by definition.

A function g that is particularly interesting is the sum of the components: $g(a, b) = a + b$. In that case, we get:

$$Q^+(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = P\big(\mathbf{x}_1 + \mathbf{x}_2 > \mathbf{y}_1 + \mathbf{y}_2\big) + \frac{1}{2}P\big(\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{y}_1 + \mathbf{y}_2\big).$$

Note that the sum gives rise to the same probabilistic relation as the arithmetic mean, and also it allows to give the following alternative expression:

$$Q^+(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = P\big(\mathbf{x}_1 - \mathbf{y}_1 > -(\mathbf{x}_2 - \mathbf{y}_2)\big) + \frac{1}{2}P\big(\mathbf{x}_1 - \mathbf{y}_1 = -(\mathbf{x}_2 - \mathbf{y}_2)\big).$$

This alternative expression allows us to give a geometric interpretation. In Fig. 1 we represent the axes $\mathbf{x}_1 - \mathbf{y}_1$ and $\mathbf{x}_2 - \mathbf{y}_2$. The dashed line represents the set $\{\mathbf{x}_1 - \mathbf{y}_1 = -(\mathbf{x}_2 - \mathbf{y}_2)$ or, equivalently, $\{\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{y}_1 + \mathbf{y}_2\}$. The gray part, that is above the dashed line, corresponds to the set $\{\mathbf{x}_1 - \mathbf{y}_1 > -(\mathbf{x}_2 - \mathbf{y}_2)$ or, equivalently, $\{\mathbf{x}_1 + \mathbf{x}_2 > \mathbf{y}_1 + \mathbf{y}_2\}$. Therefore, $Q^+(\vec{\mathbf{x}}, \vec{\mathbf{y}})$ is the probability of the gray region plus half of the probability of the dashed line.

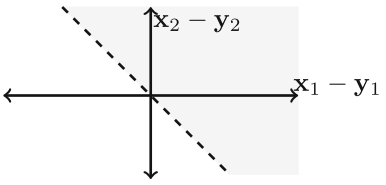


Fig. 1. Graphical representation of the probabilistic relation Q^+.

By using the winning probability Q^g we can define another extension of statistical preference to the bivariate framework.

Definition 5. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and a strictly increasing function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we say that $\vec{\mathbf{x}}$ is Q^g-statistically preferred to $\vec{\mathbf{y}}$, denoted as $\vec{\mathbf{x}} \succ^g_{\mathrm{SP}} \vec{\mathbf{y}}$, if $Q^g(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > \frac{1}{2}$.

From this definition, we get the following result.

Proposition 5. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and a strictly increasing function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the following conditions for the definition of $\vec{\mathbf{x}} \succ^g_{\mathrm{SP}} \vec{\mathbf{y}}$ are equivalent:

1. $Q^g(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > \frac{1}{2}$.
2. $Q^g(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > Q^g(\vec{\mathbf{y}}, \vec{\mathbf{x}})$.
3. $P\big(g(\vec{\mathbf{x}}) > g(\vec{\mathbf{y}})\big) > P\big(g(\vec{\mathbf{y}}) > g(\vec{\mathbf{x}})\big)$.
4. $P\big(g(\vec{\mathbf{x}}) \geq g(\vec{\mathbf{y}})\big) > P\big(g(\vec{\mathbf{y}}) \geq g(\vec{\mathbf{x}})\big)$.

3.3 Joint Winning Probability for Random Vectors

The third approach is based on using each component of both random vectors at the same time. For this aim, we introduce the joint winning probability.

Definition 6. The joint winning probability of $\vec{\mathbf{x}}$ over $\vec{\mathbf{y}}$ is the function $Q^2 : \mathcal{A}_2 \times \mathcal{A}_2 \to [0,1]$ given, for each $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$, by $Q^2(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = P\big(\mathbf{x}_1 > \mathbf{y}_1 \wedge \mathbf{x}_2 > \mathbf{y}_2\big)$.

The advantage of the function Q^2 is that it uses the joint distribution of $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$, while Q^M only uses the joint distribution of $\mathbf{x}_i$ and $\mathbf{y}_i$ and Q^g uses the joint distribution of the aggregated $g(\vec{\mathbf{x}})$ and $g(\vec{\mathbf{y}})$. Therefore, it should be more informative. On the contrary, its disadvantage is that it is no longer a probabilistic relation.

From the joint winning probability we can extend statistical preference in different ways.

Definition 7. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$, we say that:

1. $\vec{\mathbf{x}}$ is strongly statistically preferred to $\vec{\mathbf{y}}$, denoted as $\vec{\mathbf{x}} \succ^{s}_{\text{SP}} \vec{\mathbf{y}}$, if $Q^2(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > 0.5$.
2. $\vec{\mathbf{x}}$ is moderately statistically preferred to $\vec{\mathbf{y}}$, denoted as $\vec{\mathbf{x}} \succ^{m}_{\text{SP}} \vec{\mathbf{y}}$, if $Q^2(\vec{\mathbf{x}}, \vec{\mathbf{y}}) > Q^2(\vec{\mathbf{y}}, \vec{\mathbf{x}})$.
3. $\vec{\mathbf{x}}$ is weakly statistically preferred to $\vec{\mathbf{y}}$, denoted as $\vec{\mathbf{x}} \succ^{w}_{\text{SP}} \vec{\mathbf{y}}$, if $Q^2(\vec{\mathbf{y}}, \vec{\mathbf{x}}) < 0.5$.

These notions can be geometrically interpreted, as shown in Fig. 2. Considering the axes $\mathbf{x}_1 - \mathbf{y}_1$ and $\mathbf{x}_2 - \mathbf{y}_2$, $\vec{\mathbf{x}} \succ^{s}_{\text{SP}} \vec{\mathbf{y}}$ holds when the probability of the first quadrant (light gray) is greater than 0.5; $\vec{\mathbf{x}} \succ^{m}_{\text{SP}} \vec{\mathbf{y}}$ holds when the probability of the first quadrant (light gray) is greater than the probability of the third quadrant (dark gray); and finally $\vec{\mathbf{x}} \succ^{w}_{\text{SP}} \vec{\mathbf{y}}$ holds when the probability of the third quadrant (dark gray) is smaller than 0.5.

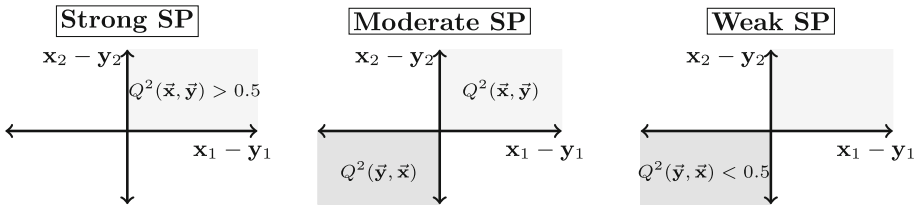


Fig. 2. Geometric explanation of the strong (left), moderate (center) and weak (right) statistical preference.

4 Connections Between the Extensions

In the previous section we have proposed several extensions of statistical preference to the comparison of (bivariate) random vectors. Along this section we compare these extensions trying to obtain connections between them.

Our first result in this direction establishes a connection between the strong, moderate and weak statistical preference, and additionally we obtain a relationship with the median similar to that in Proposition 2.

Theorem 1. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$, the following implications hold:

1. $\vec{\mathbf{x}} \succ^{s}_{\mathrm{SP}} \vec{\mathbf{y}} \;\Rightarrow\; \inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) > \vec{0} \;\Rightarrow\; \vec{\mathbf{x}} \succ^{m}_{\mathrm{SP}} \vec{\mathbf{y}} \;\Rightarrow\; \vec{\mathbf{x}} \succ^{w}_{\mathrm{SP}} \vec{\mathbf{y}}$.
2. $\vec{\mathbf{x}} \succ^{m}_{\mathrm{SP}} \vec{\mathbf{y}} \;\Rightarrow\; \inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) \not< \vec{0}$.
3. $\inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) > \vec{0} \;\Rightarrow\; \inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) \not\leq \vec{0} \;\Rightarrow\; \vec{\mathbf{x}} \succ^{w}_{\mathrm{SP}} \vec{\mathbf{y}}$.

Let us clarify that the median in the previous result is computed componentwisely. As well, $\inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) > \vec{0}$ means that the median at both components of $\vec{\mathbf{x}} - \vec{\mathbf{y}}$ is strictly positive. Similarly, $\inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) \not< \vec{0}$ (resp. $\inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) \not\leq \vec{0}$) means that at least one of the medians is non-negative (resp., strictly positive).

Next, we analyse the connection between strong and weak statistical preference and Q^M-statistical preference.

Theorem 2. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and an idempotent and strictly increasing function $M : [0,1] \times [0,1] \to [0,1]$, the following implications hold:

$$\begin{array}{ccc} \inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) > \vec{0} \Rightarrow \vec{\mathbf{x}} \succ^{M}_{\mathrm{SP}} \vec{\mathbf{y}} & \Rightarrow & \inf \mathrm{Me}(\vec{\mathbf{x}} - \vec{\mathbf{y}}) \not< \vec{0}. \\ \Downarrow & & \\ \vec{\mathbf{x}} \succ^{w}_{\mathrm{SP}} \vec{\mathbf{y}} & & \end{array}$$

Note that the previous result holds in particular for the arithmetic mean ($M = \mathrm{am}$), even if the condition in Proposition 3 is not needed.

To conclude this study, we analyse the connection between strong and weak statistical preference and Q^g-statistical preference.

Theorem 3. Given $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{A}_2$ and a strictly increasing function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the following implications hold: $\vec{\mathbf{x}} \succ^{s}_{\mathrm{SP}} \vec{\mathbf{y}} \;\Rightarrow\; \vec{\mathbf{x}} \succ^{g}_{\mathrm{SP}} \vec{\mathbf{y}} \;\Rightarrow \vec{\mathbf{x}} \succ^{w}_{\mathrm{SP}} \vec{\mathbf{y}}$.

Theorems 1–3 establish the connections between the extensions of statistical preference given in the previous section, and as well it analyses the connection with the median of the difference of the random vector. It is worth noting that strong-statistical preference is the strongest condition because it implies all the other notions. All these connections are illustrated in Fig. 3, where an arrow means that the parent implies the children, and missed arrows means that there is not a general connection between the concepts. Note that no additional implication between the different notions exists.

5 Conclusions and Further Research

This paper presents different ways of extending statistical preference for the comparison of (bivariate) random vectors. Our proposals align with three different philosophies: (i) aggregating the winning probabilities of each component; (ii) computing the winning probability of the aggregated random vectors; or (iii) directly comparing the joint distribution of both random vectors. Even though we have restricted ourselves to bivariate random vectors, the extension of the present work to the multivariate framework seems to be simple.

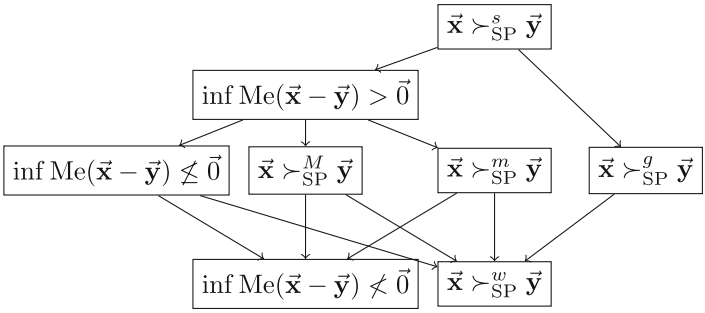


Fig. 3. Summary of the connections between the extensions of statistical preference.

These proposals have pros and cons, and a deep study on the topic is needed. Moreover, we should also investigate the connection between stochastic dominance and the new proposals. As well, the dependence structure should be taken into account, considering not only the dependence structure between the components of each random vector but also between both random vectors. Finally, it is interesting to investigate the behaviour of the proposals for well-known families of probability distributions.

Acknowledgements. This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "FEDER/UE".

References

1. Belzunce, F., Martínez-Riquelme, C., Mulero, J.: An Introduction to Stochastic Orders. Academic Press (2016)
2. Boland, P.J., Singh, H., Cukic, B.: The stochastic precedence ordering with applications in sampling and testing. J. Appl. Probab. **41**, 73–82 (2004)
3. De Meyer, H., De Baets, B., De Schuymer, B.: On the transitivity of the comonotonic and countermonotonic comparison of random variables. J. Multivar. Anal. **98**, 177–193 (2007)
4. De Schuymer, B., De Meyer, H., De Baets, B: A fuzzy approach to stochastic dominance of random variables. In: Lecture Notes in Artificial Intelligence, Vol. 2715, pp. 253–260. Springer (2003)
5. De Schuymer, B., De Meyer, H., De Baets, B., Jenei, S.: On the cycle-transitivity of the dice model. Theory Decis. **54**, 261–285 (2003)
6. Kieser, M., Friede, T., Gondan, M.: Assessment of statistical significance and clinical relevance. Stat. Med. **32**(10), 1707–1719 (2012)
7. Levy, H.: Stochastic Dominance. Kluwer Academic Publishers (1998)
8. Mann, H.B., Withney, D.R: On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18(1), 50–60 (1947)
9. Montes, I., Martinetti, D., Díaz, S., Montes, S.: Interpretation of statistical preference in terms of location parameters. INFOR:Inf. Syst. Oper. Res. **53**, 1–12 (2015)
10. Montes, I., Montes, S.: Stochastic dominance and statistical preference for random variables coupled by an Archimedean copula or by the Fréchet-Hoeffding upper bound. J. Multivar. Anal. **143**, 275–298 (2016)

11. Montes, I., Montes, S., De Baets, B.: Multivariate winning probabilities. Fuzzy Sets Syst. **362**, 129–143 (2019)
12. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risk. Wiley (2002)
13. Nelsen, R.: An Introduction to Copulas. Springer, New York (1999)
14. Vargha, A., Delaney, H.D.: The Kruskal-Wallis test and stochastic homogeneity. J. Educ. Behav. Stat. **23**(2), 170–192 (1998)

New Decision Rules for Fuzzy Statistical Inferences

Julien Rosset(✉) and Laurent Donzé

Department of Informatics, University of Fribourg, ASAM Group, Fribourg, Switzerland
{julien.rosset,laurent.donze}@unifr.ch

Abstract. We profit from the concept of the fuzzy confidence interval to implement two decision rules based on two new measures called levels of acceptance and rejection, to perform fuzzy hypothesis testing. The procedure can make a decision when both the data and hypotheses are fuzzy. Through a numerical application, we illustrate our proposed method. We compare the levels of acceptance and rejection of different fuzzy null hypotheses at a 5% significant level for a confidence interval at a 95% confidence level.

Keywords: Fuzzy inference · Fuzzy bootstrapped sample · Fuzzy confidence interval · Fuzzy test on the mean

1 Introduction

Statisticians working with fuzzy data have tried for decades to handle fuzzy inferences properly and have already worked on extending the classical inference concepts to the fuzzy world with more or less success. Nowadays, one disposes of several methods specially implemented to perform fuzzy inference (see, for example, Berkachy and Donzé ([2–5]) and Berkachy [1]). But, one major difficulty of the fuzzy inferences remains. Indeed, one cannot adequately treat the fuzzy distribution the fuzzy test statistic should follow. In this context, a straight comparison with the classical approach is sometimes not pertinent and could lead us in false directions.

In the following, we offer two new rules of decision to test a parameter θ when both fuzzy data and hypotheses are fuzzy. Our method is general and allows fuzzy inference at a given significant level. The procedure consists essentially of calculating a fuzzy confidence interval at a given confidence level for the considered parameter and comparing it to the fuzzy null hypotheses. Evidently, the fuzzy interval depends on the fuzziness of the data, which in turn influences the distribution of the test statistic. To manage this problem, Rosset and Donzé [9] recently proposed using bootstrapped techniques to construct fuzzy confidence intervals. The advantages of our proposed method are that it is intuitive in the sense that the results are easy to interpret and are not heavy computationally.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 405–412, 2024.
https://doi.org/10.1007/978-3-031-65993-5_50

The following illustrates our fuzzy testing approach with a numerical application. We thoroughly discuss and compare the results for several sets of hypotheses, which also gives us the opportunity to show how to interpret the obtained results.

In Sect. 2, we comment on recent developments and summarise the notation used throughout the study. We develop, proof and explain our new fuzzy test in Sect. 3. Then, an empirical application is given in Sect. 4. Finally, Sect. 5 concludes our work.

2 Recent Developments and Notation

In the past decades, many researchers explored the fuzzy extension of a classical confidence interval with the aim of using the confidence interval to make statistical inferences. Several fuzzy statistical testing procedures are based on this principle. Regarding the generalisation of a confidence interval to the fuzzy world, we can mention the work of Chachi and Taheri [6], Kahraman, Oztaysi and Onar [7], Berkachy [1] and Berkachy and Donzé [5]. Chachi and Taheri [6] studied the construction of fuzzy confidence intervals for fuzzy parameters based on normal fuzzy random variables and then performed tests for the mean. The restriction to normal fuzzy random variables made it unsuitable for most applications since this assumption is rarely verified. Kahraman, Oztaysi and Onar [7] developed interval-valued intuitionistic fuzzy confidence intervals (IVIF) to test population means. Despite its interesting method, the procedure remains limited since it is built upon using a crisp standard deviation and student t-distribution. The method proposed by Berkachy and Donzé [5] is overcoming this problem by considering bootstrapping techniques to construct fuzzy confidence intervals and performing a fuzzy testing procedure via the likelihood ratio. Since the concept of fuzzy likelihood ratio is difficult to generalise, we were interested in developing a testing procedure that maintains the general construction of fuzzy confidence intervals through Bootstrap while overcoming the limitation of the fuzzy likelihood ratio method to be applicable to any parameter.

In the following, we will define a fuzzy number by $\tilde{x}$. We write by $\mu_{\tilde{x}}(\cdot)$, the membership function. We consider also the α-cuts of $\tilde{x}$ denoted by $\tilde{x}^{\alpha}$ or by its equivalent in interval form by $[x^{L,\alpha}, x^{R,\alpha}]$. In practice, triangular fuzzy numbers are often used. We denote them by a triplet $\tilde{x} = (x^L, x, x^R)$ with $x^L \leq x \leq x^R \in \mathbb{R}$. We denote by X a random variable and by $\tilde{X}$ a fuzzy random variable in the epistemic sense. At last, let us designate by $\complement$ the complement of a set and by $\neg$ the negation of a preposition.

3 Fuzzy Inference for a Parameter θ

Let θ be an unknown distribution parameter, and let us draw a random sample $X_1, \ldots, X_n$ from this distribution. Assuming fuzzy modelling of the data, we can write as $\tilde{X}_1, \ldots, \tilde{X}_n$, the fuzzy equivalent of the random sample, and as $\mathbf{X} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ the fuzzy vector with its membership function $\mu_{\tilde{X}}$, such that

$\mu_{\tilde{X}} : \mathbb{R}^n \to [0,1]^n$. On the other hand, let $\phi : \mathbb{R}^n \to \mathbb{R}$ a real-valued function, and assume that ϕ is the appropriate test function for the parameter θ. In the case of fuzzy data, this function is applied to a fuzzy vector, $\phi(\underset{\sim}{\mathbf{X}}) = \phi(\tilde{X}_1, \ldots, \tilde{X}_n)$. The latter gives a fuzzy number, which we note simply by $\tilde{\phi}$ and designate its α-cuts by $\tilde{\phi}^\alpha$. Viertl [11] gives a complete definition of a fuzzy test statistic. Let us define a fuzzy test in the sense of Kruse and Meyer [8]. See also Berkachy [1, p. 122].

Definition 1. (Fuzzy Test) Consider a fuzzy random vector $\tilde{X}_1, ..., \tilde{X}_n$ related to the original $X_1, ..., X_n$ in the space Ω. Let δ be a particular value of the interval $[0,1]$. A function $\tilde{\phi}$ is called a fuzzy test for the null H_0 and the alternative H_1 hypotheses at the significance level δ if

$$\sup_{\alpha\in[0,1]} P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^\alpha\left(\tilde{X}_1(\omega_1), ..., \tilde{X}_n(\omega_n)\right) \in \{1\}\right\} \leq \delta \qquad (1)$$

where $\tilde{\phi}^\alpha$ is the α-cut of $\tilde{\phi}$ and P_{H_0} means that the probability is evaluated under the condition of the null hypothesis, i.e. $\tilde{\phi}$ is constructed to test H_0.

Adopting the procedure described, for example, in Berkachy [1], we can propose new decision rules for a test at significance level δ with fuzzy data and fuzzy hypotheses. Recall that a fuzzy hypothesis is the transcription of a crisp hypothesis into a fuzzy number, generally a triangular one. The fuzzy null and alternative hypotheses are respectively noted by $\tilde{H}_0$ and $\tilde{H}_1$.

Let us first assume that $\tilde{\sqcap}$ is a fuzzy confidence interval for the parameter θ at the confidence level $1 - \delta$. We explain below how to build such a confidence interval.

Let $\tilde{\phi}^{\text{rej}}$ and $\tilde{\phi}^{\text{acc}}$, two fuzzy statistical tests on $\mathbb{F}_c^*(\mathbb{R})$, the class of the non-empty compact, convex and normal fuzzy sets on $\mathbb{R}$, such that

$$\tilde{\phi}^{\text{rej}} : (\mathbb{F}_c^*(\mathbb{R}))^n \to [0,1], \qquad \tilde{\phi}^{\text{acc}} : (\mathbb{F}_c^*(\mathbb{R}))^n \to [0,1].$$

For a given α-cut, we assume the following functions:

$$\tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1, ..., \tilde{X}_n) = \begin{cases} 0 & \text{if} \quad \tilde{H}_0^\alpha \subseteq \tilde{\sqcap}^\alpha \quad \text{or if} \quad \alpha = 1, \\ 1 & \text{if} \quad \tilde{H}_0^\alpha \subseteq (\tilde{\sqcap}^\alpha)^{\complement}, \\ b_\alpha & \text{if} \quad \left(\tilde{H}_0^\alpha \cap (\tilde{\sqcap}^\alpha)^{\complement} \neq \emptyset \quad \text{and } \neg(\tilde{H}_0^\alpha \supset \tilde{\sqcap}^\alpha) \text{ and } \alpha < 1\right), \\ c_\alpha & \text{if } \left(\tilde{H}_0^\alpha \cap (\tilde{\sqcap}^\alpha)^{\complement} \neq \emptyset \quad \text{and } \tilde{H}_0^\alpha \supset \tilde{\sqcap}^\alpha \quad \text{and} \quad \alpha < 1\right), \\ 0 & \text{otherwise}, \end{cases} \qquad (2)$$

$$\tilde{\phi}^{\text{acc},\alpha}(\tilde{X}_1, ..., \tilde{X}_n) = \begin{cases} 0 & \text{if} \quad \left(\tilde{H}_0^\alpha \subseteq (\tilde{\sqcap}^\alpha)^{\complement} \quad \text{or} \quad \alpha = 1\right), \\ 1 & \text{if} \quad \tilde{H}_0^\alpha \subseteq \tilde{\sqcap}^\alpha, \\ d_\alpha & \text{if} \quad \left(\tilde{H}_0^\alpha \cap \tilde{\sqcap}^\alpha \neq \emptyset \quad \text{and} \quad \alpha < 1\right), \\ 0 & \text{otherwise}, \end{cases} \qquad (3)$$

where

$$b_\alpha = \left(\frac{r_R - r_L}{H_0^{R,\alpha} - H_0^{L,\alpha}}\right), c_\alpha = \left(\frac{r_R - r_L - \left(\sqcap^{R,\alpha} - \sqcap^{L,\alpha}\right)}{H_0^{R,\alpha} - H_0^{L,\alpha}}\right) \text{ and } d_\alpha = \left(\frac{a_R - a_L}{H_0^{R,\alpha} - H_0^{L,\alpha}}\right).$$

When $\alpha = 1$, $H_0^{R,\alpha} - H_0^{L,\alpha} = 0$, we cannot evaluate b_α, c_α, d_α, hence we have to define a specific rule. The quantities a_L, a_R, r_L and r_R are given by:

$$\begin{aligned} a_L &= \inf(\tilde{H}_0^\alpha \cap \tilde{\sqcap}^\alpha), & a_R &= \sup(\tilde{H}_0^\alpha \cap \tilde{\sqcap}^\alpha), \\ r_L &= \inf(\tilde{H}_0^\alpha \cap (\tilde{\sqcap}^\alpha)^{\complement}), & r_R &= \sup(\tilde{H}_0^\alpha \cap (\tilde{\sqcap}^\alpha)^{\complement}). \end{aligned}$$

Proposition 1. *$\tilde{\phi}^{rej}$ and $\tilde{\phi}^{acc}$ are fuzzy tests.*

Similar as in Berkachy [1], a proof of Prop. (1) is given as follows:

Proof. Let us examine the fuzzy test $\tilde{\phi}^{\text{rej}}$. We must verify that the test conforms to Def. 1. Consider the two-sided confidence interval $\tilde{\sqcap}$ of a parameter θ at the significance level δ. By definition, for every value of θ at the null hypothesis H_0,

$$P_{H_0}(\sqcap^{L,\alpha} \leq \theta \leq \sqcap^{R,\alpha}) \geq 1 - \delta \quad \forall \alpha \in [0,1].$$

On the other hand, we can also see that

$$P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1(\omega_1), \ldots, \tilde{X}_n(\omega_n)) \in \{1\}\right\}$$

is equivalent to

$$1 - P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1(\omega_1), ..., \tilde{X}_n(\omega_n)) \in \left\{\{0\}, \{b_\alpha\}, \{c_\alpha\}\right\}\right\}.$$

Moreover, by construction, we have that

$$P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1(\omega_1), ..., \tilde{X}_n(\omega_n)) \in \left\{\{0\}, \{b_\alpha\}, \{c_\alpha\}\right\}\right\}$$

is equal to $P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \theta \in \tilde{\sqcap}^\alpha(\omega_1, \ldots, \omega_n)\right\}$. Thus, we can write that

$$\begin{aligned} &P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \theta \in \tilde{\sqcap}^\alpha(\omega_1, \ldots, \omega_n)\right\} \\ &= P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \sqcap^{L,\alpha}(\omega_1, \ldots, \omega_n) \leq \theta \leq \sqcap^{R,\alpha}(\omega_1, \ldots, \omega_n)\right\} \\ &\geq 1 - \delta. \end{aligned}$$

Hence,

$$P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1(\omega_1), \ldots, \tilde{X}_n(\omega_n)) \in \{1\}\right\} \leq \delta,$$

which is true for all α-cuts. Therefore,

$$\sup_{\alpha \in [0,1]} P_{H_0}\left\{\omega_i \in \Omega, i = 1, \ldots, n : \tilde{\phi}^{\text{rej},\alpha}(\tilde{X}_1(\omega_1), \ldots, \tilde{X}_n(\omega_n)) \in \{1\}\right\} \leq \delta,$$

and $\tilde{\phi}^{rej}$ is a fuzzy test. The proof is exactly the same for $\tilde{\phi}^{\text{acc}}$. □

The above constructed fuzzy tests $\tilde{\phi}^{rej}$ and $\tilde{\phi}^{\text{acc}}$ permit us to propose a set of two new decision rules given in the following definition:

Definition 2. (*Decision rules*) Let be

$$L^{\text{rej}} = \int_{[0,1]} \tilde{\phi}^{\text{rej},\alpha} \mathrm{d}\alpha \qquad \text{and} \qquad L^{\text{acc}} = \int_{[0,1]} \tilde{\phi}^{\text{acc},\alpha} \mathrm{d}\alpha. \tag{4}$$

L^{rej} and L^{acc} are called, respectively, the levels of rejection and acceptance (not rejection) of the (fuzzy) null hypothesis $\tilde{H}_0$.

1. If $L^{\text{rej}} = 1$, $\tilde{H}_0$ is rejected. In this case $L^{\text{acc}} = 0$;
2. If $L^{\text{rej}} = 0$, $\tilde{H}_0$ is not rejected. In this case $L^{\text{acc}} = 1$;
3. If $0 < L^{\text{rej}} < 1$, the test shows how strong $\tilde{H}_0$ is rejected. The higher L^{rej} is, the stronger the test rejects $\tilde{H}_0$;
4. If $0 < L^{\text{acc}} < 1$, the test shows how strong $\tilde{H}_0$ is not rejected. The higher L^{acc} is, the stronger the test doesn't reject $\tilde{H}_0$;
5. If $L^{\text{rej}} > L^{\text{acc}}$, $\tilde{H}_0$ tends to be rejected;
6. If $L^{\text{rej}} < L^{\text{acc}}$, $\tilde{H}_0$ tends to be not rejected;
7. If $L^{\text{rej}} = L^{\text{acc}}$, no decision can be taken.

The idea for this fuzzy inference is to construct a fuzzy decision rule based on the two ratios of the area of the considered fuzzy hypothesis inside the rejection domain, noted by $\neg\tilde{\sqcap}$ or/and inside the acceptance domain, given by $\tilde{\sqcap}$, divided by the area of $\tilde{H}_0$ and to compare them.

The fuzzy tests $\tilde{\phi}^{rej}$ and $\tilde{\phi}^{\text{acc}}$ depend on a fuzzy confidence interval $\tilde{\sqcap}$ at a given confidence level $1-\delta$. The question of the construction of such a fuzzy confidence interval becomes a central point of the inference procedure. In Rosset and Donzé [10], we propose constructing a fuzzy confidence interval using bootstrap techniques. Our method is attractive because no specific (fuzzy) distribution should be assumed. Let us recall our method.

Let be $\tilde{t}$ a fuzzy estimator for the fuzzy parameter $\tilde{\theta}$. By using bootstrap techniques, we can estimate k new fuzzy parameters $\tilde{t^*}_1, \ldots, \tilde{t^*}_k$ and build an estimated empirical fuzzy distribution $\tilde{\hat{F}}$ with the $k+1$ estimates ("augmented distribution"). Left and right parts of a fuzzy confidence interval at a given significance level γ for each α-cut are given by:

$$\hat{\theta}^{L,\alpha}_{\gamma/2} = t^{L,\alpha} - (t^{*L,\alpha}_{(k+1)(1-\gamma/2)} - t^{L,\alpha}) = 2t^{L,\alpha} - t^{*L,\alpha}_{(k+1)(1-\gamma/2)} \tag{5}$$

$$\hat{\theta}^{R,\alpha}_{1-\gamma/2} = t^{R,\alpha} - (t^{*R,\alpha}_{(k+1)(\gamma/2)} - t^{R,\alpha}) = 2t^{R,\alpha} - t^{*R,\alpha}_{(k+1)(\gamma/2)}. \tag{6}$$

where, for the left α-cut, $t^{*L,\alpha}_{(k+1)(1-\gamma/2)}$ is the $(1-\gamma/2)$-quantile, and for the right α-cut, $t^{*R,\alpha}_{(k+1)(\gamma/2)}$ is the $\gamma/2$-quantile of the augmented distribution. The bootstrapped fuzzy confidence interval at confidence level $1-\gamma$ is given by $\tilde{\sqcap}^{\alpha} = [\sqcap^{L,\alpha}, \sqcap^{R,\alpha}] = [\hat{\theta}^{L,\alpha}_{\gamma/2}, \hat{\theta}^{R,\alpha}_{1-\gamma/2}]$.

4 Application

We intend to test the mean of a distribution at a 5% significant level. We consider the observations from Berkachy and Donzé [5] given in Table 1. The data is supposed to be fuzzy. Table 1 shows how the data has been fuzzified. The crisp mean is 2.8, and the fuzzy one is $\bar{\tilde{x}} = (1.8, 2.8, 3.8)$. In the first step, we generate a fuzzy confidence interval at a 95% confidence level with the method described above. We obtain $(0.9333, 1.9333, 3.6, 4.6)$. The two fuzzy tests can be computed after postulating a fuzzy null hypothesis. We offer in Table 2 the levels of acceptance L^{acc} and rejection L^{rej} for 13 different null fuzzy hypotheses.

Table 1. Fuzzy observations from Berkachy and Donzé [5]

x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$
4	$(3,4,5)$	3	$(2,3,4)$	3	$(2,3,4)$	5	$(4,5,6)$	3	$(2,3,4)$
1	$(0,1,2)$	2	$(1,2,3)$	2	$(1,2,3)$	2	$(1,2,3)$	3	$(2,3,4)$

Note $\bar{x} = 2.8$, $\bar{\tilde{x}} = (1.8, 2.8, 3.8)$

4.1 Discussion: Acceptance and Rejection Levels

Looking at Table 2, which shows the levels of acceptance and rejection of given fuzzy null hypotheses, we first notice that when the fuzzy null hypothesis is outside the fuzzy confidence interval, its level of rejection is 1 and its level of acceptance is 0 and conversely, when the fuzzy null hypothesis is inside the fuzzy confidence interval, its level of rejection is 0, and its level of acceptance is 1. This result is not surprising and is in accordance with crisp classical tests. Let us now discuss the cases where the null hypothesis overlaps the fuzzy confidence interval. We see that when we consider a fuzzy null hypothesis larger than the fuzzy confidence interval, given by $\tilde{H}_0 = (0, 3.8, 6)$ in Table 2 (Test 11), we find a level of acceptance of 0.68 and a level of rejection of 0.43. These levels describe how much $\tilde{H}_0$ is inside, respectively, outside the fuzzy confidence interval. In this case, we would tend not to reject the fuzzy null hypothesis. However, if we take another larger $\tilde{H}_0 = (0.5, 4.5, 8)$ (Test 12), this time, a more significant portion is out of the fuzzy confidence interval and is inside its complementary, thus telling us we should tend to reject the null hypothesis more often than accepting it. It is pretty natural to find such results. The levels of acceptance and rejection reflect the portion of the null hypothesis contained in the confidence interval, respectively, included in its complementary. Lastly, we can also see that when we use crispier but still fuzzy null hypotheses (Tests 2,3,4,6,8,9,10), the results tend to be more and more binary. This comes from the fact that the thinner $\tilde{H}_0$ is, the easier it is for it to be outside or inside the confidence interval. Notice that when a thin $\tilde{H}_0$ (Tests 5,7) is outside the fuzzy confidence interval's core but still inside its fuzzy left and right bounds, we again get non-binary rejection

and acceptance levels. This is because $\tilde{H}_0$ is simultaneously inside the fuzzy confidence interval and its complementary.

Table 2. Fuzzy Hypothesis testing at a 5% significant level

Tests	$\tilde{H}_0$	acceptance	rejection
1	(−0.3, −0.2, −0.1)	0	1
2	(0, 0.5, 0.8)	0	1
3	(0.8, 0.85, 0.9)	0	1
4	(0.95, 1, 1.05)	0.12	0.99
5	(1, 1.5, 2)	0.75	0.58
6	(1.8, 1.9,2)	0.99	0.08
7	(1.7, 1.75, 1.78)	0.96	0.34
8	(2, 2.5, 3.5)	1	0
9	(2.4, 2.45, 2.5)	1	0
10	(2.6, 2.8, 3.2)	1	0
11	(0, 3.8, 6)	0.68	0.43
12	(0.5, 4.5, 8)	0.44	0.62
13	(7, 8, 9)	0	1

This Table shows the levels of acceptance and rejection of different fuzzy null hypotheses. The fuzzy trapezoidal number $(0.933, 1.933, 3.6, 4.6)$ is a fuzzy confidence interval at a 95% confidence level for the distribution's mean

5 Conclusion

Our work introduced a fuzzy statistical testing procedure based on precedent works. Our new method revolves around constructing a fuzzy confidence interval for the parameter of interest and comparing it to a fuzzy null hypothesis. More precisely, we compare the ratios of how much of the null hypothesis is contained respectively in the confidence interval or its complementary. We then illustrated our testing procedure through a numerical application and discussed the results. The crisp nature of the results of our testing method, namely the levels of acceptance and rejection, allows one to interpret them easily. Moreover, the underlying simplicity of the method is of great advantage. Indeed, it is relatively easy to implement and is not intensive computation-wise. Furthermore, its general aspect should also be mentioned. Indeed, to perform statistical testing, one just calculates a fuzzy confidence interval for one's parameter of interest to perform statistical testing. Lastly, let us again mention that it generalises the classical testing approach. Indeed, the crispier the null hypothesis is, the more binary the results are. In the future, for further research, we will be interested in investigating the behaviour of our testing procedure for different parameters.

References

1. Berkachy, R.: The Signed Distance Measure in Fuzzy Statistical Analysis. Some theoretical, empirical and programming advances. PhD thesis University of Fribourg (2020). https://folia.unifr.ch/unifr/documents/308986
2. Berkachy, R., Donzé, L.: Testing fuzzy hypotheses with fuzzy data and defuzzification of the fuzzy p-value by the signed distance method. In: Proceedings of the 9th International Joint Conference on Computational Intelligence, pp. 255–264 (2017). https://doi.org/10.5220/0006500602550264
3. Berkachy, R., Donzé, L.: Fuzzy confidence interval estimation by likelihood ratio. In: Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019). Atlantis Studies in Uncertainty Modelling Vol. 1., pp. 150–157. Atlantis Press (2019a). https://doi.org/10.2991/eusflat-19.2019.23
4. Berkachy, R., Donzé, L.: Testing hypotheses by fuzzy methods: a comparison with the classical approach. In: Meier, A., Portmann, E., Terán, L. (Eds.) Applying Fuzzy Logic for the Digital Economy and Society, pp. 1–22. Fuzzy Management Methods. Springer, Cham (2019b). https://doi.org/10.1007/978-3-030-03368-2_1
5. Berkachy, R., Donzé, L.: Fuzzy Confidence Intervals by the Likelihood Ratio with Bootstrapped Distribution. In: Proceedings of the 12th International Joint Conference on Computational Intelligence (IJCCI 2020) - Volume 1, , pp. 231–242. SciTePress (2020). https://doi.org/10.5220/0010023602310242
6. Chachi, J., Taheri, S.: Fuzzy confidence intervals for mean of Gaussian fuzzy random variables. Expert Syst. Appl. 38(5), 5240–5244 (2011). https://doi.org/10.1016/j.eswa.2010.10.034
7. Kahraman, C., Oztaysi, B., Onar, S.C. Interval-Valued Intuitionistic Fuzzy Confidence Intervals. J. Intell. Syst. **28**(2), 307–319 (2019). https://doi.org/10.1515/jisys-2017-0139
8. Kruse, R., Meyer, K.D.: Statistics with Vague Data. Theory and Decision Library, Series B: Mathematical and Statistical Methods. D. Reidel Publishing Company (1987). https://doi.org/10.1007/978-94-009-3943-1
9. Rosset, J., Donzé, L.: Fuzzy least squares and fuzzy orthogonal least squares linear regressions. In: Proceedings of the 15th International Joint Conference on Computational Intelligence - FCTA, pp. 359–368. SciTePress (2023). https://doi.org/10.5220/0012182700003595
10. Rosset, J., Donzé, L.: Fuzzy confidence intervals by bootstrapped fuzzy distributions. In: Proceedings of the 20th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2024, Lisboa, July 22–26, 2024) (2024)
11. Viertl, R.: Statistical Methods for Fuzzy Data. John Wiley & Sons (2011). https://doi.org/10.1002/9780470974414

Learning in Reproducing Kernel Hilbert Spaces for Orbits of Iterated Function Systems

Priyanka Roy[1,2(✉)] and Susanne Saminger-Platz[2]

[1] Software Competence Center Hagenberg, 4232 Hagenberg, Austria
priyanka.roy@scch.at, priyanka.roy@jku.at
[2] Institute for Mathematical Methods in Medicine and Data-Based Modeling, Johannes Kepler University Linz, 4020 Linz, Austria
susanne.saminger-platz@jku.at

Abstract. One of the problems in learning theory is to approximate a function f that underlies the relationship between (x, y), i.e., $y = f(x)$ based on sample points $(x_t, y_t)_{t=1}^n$. Given the sample points, the function can be approximated in Reproducing Kernel Hilbert Spaces through various learning algorithms. However, it is usually customary to consider the sampling nature as independent and identically distributed (i.i.d.) in the context of learning theory. We leverage the i.i.d. assumption by considering an input sample trajectory $(x_t)_{t\in\mathbb{N}}$ obtained via an Iterated Function System that is a particular Markov Chain, with $(y_t)_{t\in\mathbb{N}}$ corresponding to an observation sequence when the model is in the corresponding state x_t. We discuss learning bounds for approximation for such a process.

Keywords: Markov chain · Copula · Approximation · Compact operator · Reproducing kernel hilbert space

1 Introduction

The classical problem of learning from independent and identically distributed sample points $(z_t = (x_t, y_t))_{t=1}^n$ drawn from a probability measure ρ on $Z = X \times Y$, aims at approximating a regression function

$$f_\rho(x) = \int_Y y d\rho(y|x),$$

from a function space where $\rho(y|x)$ denotes the conditional measure. It can be shown that the regression function f_ρ minimizes the following mean square error

$$\mathcal{E}(f) = \int_{X\times Y} (f(x) - y)^2 d\rho(z).$$

Of the many interests in the field of Learning Theory, many contributions have been in the direction of obtaining error bounds for the function approximated

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 413–420, 2024.
https://doi.org/10.1007/978-3-031-65993-5_51

based on i.i.d. samples $z_t = (x_t, y_t)$ due to some learning algorithm (see, e.g., [3, 9]).

A large volume of literature explores the hypothesis space $\mathcal{H}$ as a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ due to some positive semi-definite kernel k for approximating functions by learning algorithms (see, e.g., [3,9,10]). Any function in $L^2_{\rho_X}$, the space of square-integrable functions with respect to the marginal probability measure ρ_X, can be approximated in $\mathcal{H}_k$ by choosing suitable kernels.

In this paper, we relax the i.i.d. assumption by expecting some dependencies in between neighboring sample items in time by considering an example of a Markov Chain whose input sample trajectory $(x_t)_{t\in\mathbb{N}}$ is obtained via an Iterated Function System (IFS) together with an observation sequence $(y_t)_{t\in\mathbb{N}}$, and investigate learning bounds for such realizations.

Since this work draws its formalization and context from various areas of mathematics, we recall some basic definitions in Sect. 2 that will be needed in the later developments. Section 3 will look at some convergence results of empirical copulas induced by orbits of IFS from [12]. Section 4 will contain some important discussions following [11] on the factorization of operators and therefore the embedding of function spaces due to integral operators. Finally, in Sect. 5 we state and discuss the main result.

2 Preliminaries

We briefly summarize some concepts relevant for our investigations.

Definition 1. (*The L^p spaces*) Let (X, Σ) be a measure space. For $p \in (1, \infty)$, we define the space

$$\mathcal{L}^p(\mu) = \{f : X \to \mathbb{R} : f \text{ measurable, and } \int_X |f|^p \in d\mu < \infty\}.$$

For $f \in \mathcal{L}^p(\mu)$, we further write

$$[f]_\sim := \{g \in \mathcal{L}^p(\mu) : \mu(\{f \neq g\}) = 0\}$$

for the μ equivalence class of f. As usual, we write $L^p(\mu) := \mathcal{L}^p(\mu)\backslash_\sim$ for the corresponding quotient space and $||\cdot||_{\mathcal{L}^p(\mu)}$ and $||\cdot||_{L^p(\mu)}$ for the semi-norm of $\mathcal{L}^p(\mu)$ and norm of $L^p(\mu)$, respectively, where for $0 < p < \infty$,

$$||f||_{\mathcal{L}^p(\mu)} = ||f||_{L^p(\mu)} = \left(\int |f|^p d\mu\right)^{\min\{\frac{1}{p},1\}}.$$

Definition 2. (*Positive definite Kernel and Mercer kernel*) A positive definite (p.d.) kernel on a set X is a function $k : X \times X \to \mathbb{R}$ that is symmetric i.e.,

$$\text{for all } (x, x') \in X^2, \quad k(x, x') = k(x', x),$$

and satisfies, for all $n \in \mathbb{N}$, all $(x_1, x_2, \cdots, x_n) \in X^n$ and all $(a_1, a_2, \cdots, a_n) \in \mathbb{R}^n$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0.$$

A kernel k on a set X is called a Mercer kernel if X is a compact metric space and $k : X \times X \to \mathbb{R}$ is a continuous positive definite kernel.

Definition 3. (*Reproducing Kernel Hilbert Space (RKHS)*) Let X be a set and $\mathcal{H} \subset \mathbb{R}^X$ be a class of functions forming a (real) Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The function $k : X \times X \to \mathbb{R}$ is called a reproducing kernel (r.k.) of $\mathcal{H}$ if

(i) $\mathcal{H}$ contains all function of the form

$$\forall x \in X, \ k_x : t \to k(x, t);$$

(ii) For every $x \in X$ and $f \in \mathcal{H}$ the reproducing property holds:

$$f(x) = \langle f, k_x \rangle_{\mathcal{H}}.$$

If a r.k. exists, then $\mathcal{H}$ is called the reproducing kernel Hilbert space (RKHS).

For more details on the theory of kernels, Mercer kernels, and RKHS, we refer the reader to [9,11].

Since throughout we will be dealing with compact, positive, self-adjoint operators, we kindly refer the reader to [8] for its definitions and further reading.

Definition 4. ([1,4,7] (*Copulas*)) A function $C : \mathbb{I}^d \to \mathbb{I}$ is a ***d*-copula** if, and only if, the following conditions hold:

(i) $C(u_1, \ldots, u_d) = 0$ whenever $u_j = 0$ for at least one index $j \in \{1, \ldots, d\}$,
(ii) when all the arguments of C are equal to 1, but possibly for the j-th one, then

$$C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$$

(iii) C is d-increasing i.e., $\forall\ ]\mathbf{a}, \mathbf{b}] \subseteq \mathbb{I}^d, V_C(]\mathbf{a}, \mathbf{b}]) := \sum_{\mathbf{v} \in \mathrm{ver}(]\mathbf{a}, \mathbf{b}])} \mathrm{sign}(\mathbf{v}) C(\mathbf{v}) \geq 0$

where

$$\mathrm{sign}(\mathbf{v}) = \begin{cases} 1, & if\ v_j = a_j \text{ for an even number of indices,} \\ -1, & if\ v_j = a_j \text{ for an odd number of indices.} \end{cases}$$

Note that every copula $C \in \mathcal{C}_d$ induces a *d-fold stochastic measure* μ_C on $(\mathbb{I}^d, \mathcal{B}(\mathbb{I})^d)$ defined on the rectangles $R =]\mathbf{a}, \mathbf{b}]$ contained in $\mathbb{I}^d$, by

$$\mu_C(R) := V_C(]\mathbf{a}, \mathbf{b}]).$$

Definition 5. ([2] (*Iterated Function System*)) Consider a state space $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$ and let $G := \{h_i |\ h_i : \mathbb{S} \to \mathbb{S}, i = 1, 2, \ldots, L\}$ with $L < \infty$. Let $\mu = \{p_i\}_{i=1}^{L}$ be a probability distribution. Let $\{f_n\}_{n=1}^{\infty}$ be i.i.d., such that $P(f_n =$

$h_i) = p_i$, for all $i \in \{1, \ldots, L\}$. Let X_0 be an $\mathbb{S}$-valued random variable independent of $\{f_n\}_{n=1}^{\infty}$. Then, the i.i.d. random iteration scheme $X_j = f_j(X_{j-1}),\ j \geq 1$ is a Markov chain with transition probability $P\colon \mathbb{S} \times \mathcal{B}(\mathbb{S}) \to \mathbb{R}$

$$P(s, A) = \sum_{r=1}^{L} p_r \mathbb{I}_A(h_r(s)).$$

When each h_i is an affine contraction, then the Markov Chain $\{X_n\}$ converges in distribution to some π. Further the limit point set of the process $\{X_n\}$ coincides with the support of the limit distribution π with probability 1.

3 Convergence of Empirical Copulas Induced by Orbits of IFS

Given a state space $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$, convergence in distribution of the trajectory of a given Markov process has certainly been of interest since the past decade (see, e.g., [5]). In this section, we shall briefly recall some results due to Trutschnig and Fernández Sánchez on such convergence results, for more details see [12].

Definition 6. ([6,12]) A transformation matrix T is a matrix with non-negative entries, for which the entries sum up to 1 and no row or column has all entries 0. Denote $\mathcal{T}$ as the class of all transformation matrices.

For a given transformation matrix $T \in \mathcal{T}$, set $\widehat{I} := \{(i, j) : t_{ij} > 0\}$ and consider the IFS $\left\{[0,1]^2, (f_{ij})_{(i,j)\in\widehat{I}}, (t_{ij})_{(i,j)\in\widehat{I}}\right\}$, whereby each affine contraction $f_{ij} : [0,1]^2 \to R_{ij}$ is given by

$$f_{ij}(x, y) = \Big(a_{i-1} + x(a_i - a_{i-1}), b_{j-1} + y(b_j - b_{j-1})\Big)$$

where a_i denotes the sum of the entries of the first i columns of T with $a_0 = 0$ and b_i denotes the sum of the entries of the first (bottom to top) j rows of T with $b_0 = 0$.

Theorem 1. *([12]) Suppose that $T \in \mathcal{T}$, let $\{\mathbb{I}^2, (\omega_l)_{l=1}^N, (p_l)_{l=1}^N\}$ denote the corresponding IFS and fix $(x_0, y_0) \in \mathbb{I}^2$. Then there exists a set $\Lambda \subseteq \Sigma_N$ with $P_T(\Lambda) = 1$, where $\Sigma_N = \{(k_i)_{i\in\mathbb{N}} : 1 \leq k_i \leq N \text{ for all } i \in \mathbb{N}\}$ denotes the code space of N symbols, such that for every $k \in \Lambda$, the sequence $(E'_n)_{n\in\mathbb{N}}$ of empirical copulas of the sample $(Y_i^{(x_0,y_0)}(k))_{i=1}^n := (\omega_{k_n} \circ \omega_{k_{n-1}} \circ \cdots \circ \omega_{k_1}(x_0, y_0))_{i=1}^n$ with transition probabilities $P((x, y), A) = \sum_{r=1}^N p_r \mathbb{I}_A(w_r(x, y))$, fulfills*

$$\lim_{n\to\infty} d_\infty((E'_n)_{n\in\mathbb{N}}, A_T^\star) = 0.$$

In other words: with probability one the empirical copula converges uniformly to the copula $A_T^\star$, where $A_T^\star$ is referred to as the *invariant copula*, such that $\mathcal{V}_T(\mu_{A^\star}) = \mu_{A^\star}$ with $\mathcal{V}_T$ as the induced operator on $\mathcal{P}([0,1])^2$ defined by

$$\mathcal{V}_T(\mu) := \sum_{(i,j)\in\widehat{I}} t_{ij} \mu^{f_{ji}},$$

where $\mu^{f_{ji}}$ is the pushforward measure.

4 Factorization and Embedding

In Sect. 3 we briefly looked at the trajectory due to some IFS and the corresponding invariant measure of the process which sheds light on the underlying process from where we obtain input realizations. But as mentioned in the introduction, learning takes place in a hypothesis space which is the reproducing kernel Hilbert space. Hence, for approximation purposes, a brief understanding on the embedding of function spaces is also necessary and for this the integral operators and its power play a key role. We recall some important results in this direction due to Steinwart and Scovel, for more details see also [3,11].

Let μ be any Borel measure on X, and L^2_μ be the Hilbert space of (equivalence classes of) square-integrable functions on X. For any function $k : X \times X \to \mathbb{R}$, let the transform $T_k : L^2_\mu \to L^2_\mu$ be

$$T_k f := \int_X k(\cdot, x) f(x) d\mu(x), \quad \text{for all } f \in L^2_\mu.$$

If k is a Mercer kernel, then T_k is a compact and bounded linear operator over $L^2(\mu)$. Additionally, T_k is self-adjoint, positive, and in fact maps to a set of continuous functions i.e., $C(X) \hookrightarrow L^2(\mu)$.

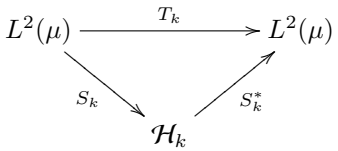
Further let X be a measurable space, μ be a measure on X, and $\mathcal{H}_k$ be a RKHS over X with measurable kernel $k : X \times X \to \mathbb{R}$. We say that $\mathcal{H}_k$ is embedded into $L^2(\mu)$ if $[f]_\sim \in L^2(\mu)$ for all $f \in \mathcal{H}_k$ and the linear operator

$$S_k^* : \mathcal{H}_k \to L^2(\mu), \quad S_k^*(f) = [f]_\sim$$

is continuous. Moreover, $\mathcal{H}_k$ is compactly embedded in $L^2(\mu)$ if S_k^* is compact and the adjoint $S_k : L^2(\mu) \to \mathcal{H}_k$ of the embedding $S_k^* : \mathcal{H}_k \to L^2(\mu)$ satisfies:

$$S_k f := \int_X k(\cdot, x') f(x') d\mu(x').$$

Thus we have the following factorization:



Note that due Mercer's theorem, for all fixed $x \in \text{supp } \mu$, we have

$$k(x, x') = \sum_{i \in I} \sigma_i \psi_i(x) \psi_i(x'), \tag{1}$$

where the convergence is uniform, with σ_i and ψ_i corresponding to the eigenvalues and eigenfunctions of T_k. Due to some additional properties on the behavior of the eigenvalues and eigenfunctions of T_k along with the non-injectivity of S_k^*, if k is continuous, then the convergence in (1) is uniform in x, x' on all

$A \times A \subset \text{supp}\ \mu \times \text{supp}\ \mu$ for which A is compact. This shows that (1) converges for all $x, x' \in \text{supp}\ \mu$. Consequently if, $\mu(X \backslash \text{supp}\ \mu) = 0$, then (1) holds on a set of the form $(X \backslash N) \times (X \backslash N)$, where $\mu(N) = 0$. We refer the reader to [11, Cor. 3.5] for further details.

5 Setting of Learning in RKHS

Definition 7. For a d-fold stochastic measure μ_C induced by a copula $C \in \mathcal{C}_d$, we define an integral operator $T_{k,\mu_C} : L^2(\mu_C) \to L^2(\mu_C)$ as

$$T_{k,\mu_C} f(\cdot) = \int_X k(\cdot, x) f(x) d\mu_C(x), \quad f \in L^2(\mu_C) \tag{2}$$

where T_{k,μ_C} is a well-defined continuous and compact operator with $L^2(\mu_C)$ being the Hilbert Space of square-integrable functions with respect to μ_C. The *regression function* f_ρ is said to satisfy the **source condition** (of order β) if

$$f_\rho = T^\beta_{k,\mu_C}(g) \text{ for some } g \in L^2(\mu_C).$$

Hence for any regression function f_ρ satisfying the source condition(of order β) we have that $f_\rho \in [H]^\beta_\sim$, where

$$[\mathcal{H}]^\beta_\sim = \{a_i \sigma_i^{\frac{\beta}{2}} [\psi_i]_\sim : (a_i) \in \ell^2(I)\}.$$

is the RKHS of T^β_{k,μ_C}, where $([\psi_i]_\sim)_{i \in I}$ consists of the eigenvectors of T_{k,μ_C} corresponding to its eigenvalues σ_i.

Hence for the invariant copula, $A^\star_T$ due to an IFS induced by a transformation matrix $T \in \mathcal{T}$, the **source condition** can be re-written as

$$f_\rho = T^\beta_{k,\mu_{A^\star_T}}(g) \text{ for some } g \in L^2(\mu_{A^\star_T}). \tag{3}$$

5.1 Error Decomposition

Let $z_t = (x_t, y_t)_{i=1}^n$ be a sample with $(x_t)_{t=1}^n$ being the state sequence generated by the IFS induced by a transformation matrix T and $(y_t)_{t=1}^n$ being a sequence of observations with ρ as the joint probability measure on $Z = \mathbb{I}^2 \times \mathbb{R}$ and regression function f_ρ. For better readability, we will abbreviate $\mu_{A^\star_T}$ simply by $\mu^\star$.

We now estimate the expectation of the error $\|f_{T+1} - f_\rho\|^2_\rho$, i.e., of $\mathbb{E}_{z_1,\ldots,z_t}(\|f_{T+1} - f_\rho\|^2_\rho)$, for which $f_{T+1} - f_\rho$ can be decomposed into the following parts:

$$f_{T+1} - f_\rho = f_{T+1} - f_{\lambda,\mu^\star} + f_{\lambda,\mu^\star} - f_\rho$$

where

$$f_{\lambda,\mu^\star} = \underset{f \in \mathcal{H}_K}{\arg\min} \left\{ \int_{\mathbb{I}^2} (f(x) - f_\rho(x))^2 d\mu^\star + \lambda \|f\|^2_K \right\}, \tag{4}$$

with $\lambda > 0$ and $\|f\|_\rho = \|f\|_{L^2(\mu^\star)}$. Further, note that $\mathbb{E}_{z_1,\ldots,z_t}$ will be denoted as $\mathbb{E}_{Z^T}$ for $T \in \mathbb{N}$.

Based on all the previous contexts discussed so far, we are now in the position to study the upper bound for the expectation of the error $\|f_{T+1} - f_\rho\|^2$. Also since $\mathbb{E}_{Z^T}[\|f_{T+1} - f_\rho\|_\rho^2]$ can be evaluated as

$$\mathbb{E}_{Z^T}[\|f_{T+1} - f_{\lambda,\mu^\star}\|_\rho^2] + \|f_{\lambda,\mu^\star} - f_\rho\|_\rho^2 + 2\mathbb{E}_{Z^T}[\langle f_{T+1} - f_{\lambda,\mu^\star}, f_{\lambda,\mu^\star} - f_\rho\rangle_\rho], \quad (5)$$

we shall analyze each term separately and provide their corresponding upper bounds as well. In our presentation we will discuss the influence of the sample based algorithm generating f_{T+1} from f_T on the determination of the bounds for the first and the last term.

The middle term in (5), however, involves the deterministic error $\|f_{\lambda,\mu^\star} - f_\rho\|_\rho$ for which the following proposition gives an analytic bound in terms of the largest eigenvalue of the operator $T_{k,\mu^\star}$.

Proposition 1. *Assume $X = \mathbb{I}^2$ and $Y = \mathbb{R}$. Consider a transformation matrix T, its induced iterated function system (IFS), invariant copula $A_T^\star$ and induced d-fold stochastic measure $\mu^\star$, i.e., $\mu^\star := \mu_{A_T^\star}$. For any Mercer kernel $k\colon X \times X \to \mathbb{R}$, the integral operator $T_{k,\mu^\star}\colon L^2(\mu^\star) \to L^2(\mu^\star)$ is defined by Eq. (2).*

Furthermore, for some $0 < \beta \leq 1$ and $(\psi_i)_{i\in I} \subset \mathcal{H}_k$ such that $([\psi_i]_\sim)_{i\in I}$ is an orthonormal system in $L_2(\mu^\star)$ with $(\sigma_i)_{i\in I} \subset (0,\infty)$ an at most countable decreasing family converging to 0 such that for all $f \in L_2(\mu^\star)$

$$T_{k,\mu^\star} f = \sum_{i\in I} \sigma_i \langle f, [\psi_i]_\sim\rangle_{L_2(\mu^\star)} [\psi_i]_\sim$$

with convergence in $L_2(\mu^\star)$ and such that $\sum_{i\in I} \sigma_i^\beta \psi_i^2(x) < \infty$, we denote the RKHS of $T_{k,\mu^\star}^\beta$ by

$$[\mathcal{H}]_\sim^\beta = \{a_i \sigma_i^{\frac{\beta}{2}} [\psi_i]_\sim : (a_i) \in \ell^2(I)\}.$$

Due to the Spectral Theorem on compact, positive and self-adjoint operator, the families $(\sigma_i)_{i\in I}$ and $(([\psi_i])_\sim)_{i\in I}$ necessarily consist of non-zero eigenvalues and the corresponding eigenvectors of $T_{k,\mu^\star}$.

Let $z_t = (x_t, y_t)_{i=1}^n$ be a sample with $(x_t)_{t=1}^n$ being the state sequence generated by the IFS and $(y_t)_{t=1}^n$ being a sequence of observations and let ρ be the joint probability measure on $Z = \mathbb{I}^2 \times \mathbb{R}$ and regression function f_ρ.

Then the following holds: If f_ρ satisfies the source condition (of order β) for some $g \in L^2(\mu^\star)$, then for any $\lambda > 0$, we have

$$\|f_{\lambda,\mu^\star} - f_\rho\|_{[H]_\sim^\beta} \leq \sigma_1^{\frac{\beta}{2}} \|g\|_{L^2(\mu^\star)}$$

with $f_{\lambda,\mu^\star}$ as in (4) and σ_1 being the largest eigenvalue of $T_{k,\mu^\star}$.

Acknowledgements. This research was carried out under the Austrian COMET program (project S3AI with FFG no. 872172, www.S3AI.at, at SCCH, www.scch.at), which is funded by the Austrian ministries BMK, BMDW, and the province of Upper Austria.

References

1. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific Publishing, Hackensack, NJ (2006)
2. Athreya, K.B., Lahiri, S.N.: Measure Theory and Probability Theory. Springer Texts in Statistics. Springer, New York (2006)
3. Cucker, F., Smale, S.: On the mathematical foundations of learning. Bull. Amer. Math. Soc. (N.S.) **39**(1), 1–49 (2002)
4. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press (2015)
5. Elton, J.H.: An ergodic theorem for iterated maps. Ergodic Theory Dynam. Syst. **7**(4), 481–488 (1987)
6. Fredricks, G.A., Nelsen, R.B., Rodríguez-Lallena, J.A.: Copulas with fractal supports. Insurance Math. Econom. **37**(1), 42–48 (2005)
7. Nelsen, R.B.: An Introduction to Copulas. Springer Series in Statistics, 2nd edn. Springer, New York (2006)
8. Simon, B.: Operator theory, Vol. 4. American Mathematical Soc. (2015)
9. Smale, S., Yao, Y.: Online learning algorithms. Found. Comput. Math. **6**(2), 145–170 (2006)
10. Smale, S., Zhou, D.-X.: Online learning with Markov sampling. Anal. Appl. **7**(01), 87–113 (2009)
11. Steinwart, I., Scovel, C.: Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs. Constr. Approx. **35**(3), 363–417 (2012)
12. Trutschnig, W., Fernández Sánchez, J.: Copulas with continuous, strictly increasing singular conditional distribution functions. J. Math. Anal. Appl. **410**(2), 1014–1027 (2014)
13. Ying, Y., Pontil, M.: Online gradient descent learning algorithms. Found. Comput. Math. **8**(5), 561–596 (2008)

A Comparative Analysis of Aggregation Rules for Coherent Lower Previsions

Juan J. Salamanca, Ignacio Montes, and Enrique Miranda(✉)

Department of Statistics and O.R., University of Oviedo,
C/Leopoldo Calvo Sotelo 18, 33007 Oviedo, Spain
{salamancajuan,imontes,mirandaenrique}@uniovi.es

Abstract. We consider the problem of aggregating information provided by experts when this information is expressed by means of coherent lower previsions. These constitute a framework general enough so as to include as particular cases not only probability measures but also the majority of models from the imprecise probability theory. Although the aggregation problem has already been tackled in the literature, our contribution provides a unified view by putting together a number of rationality criteria and aggregation rules studied in different papers. Specifically, we consider five aggregation rules, twelve rationality criteria and provide a detailed analysis of the properties satisfied by each rule.

Keywords: Probabilistic opinion pooling · Aggregation rules · Coherent lower previsions · Natural extension · Credal sets

1 Introduction

The mathematical modelling of uncertainty suffers quite frequently from the presence of vague or conflicting information, as well as the existence of missing data. To deal with these situations, a number of models have been proposed in the literature, and they are nowadays usually encompassed with the term *imprecise probabilities* [8]. While imprecise probabilities allow an expert for greater flexibility and robustness in the modelling, as well as an arguable more realistic account of the available information, their use also introduces a number of challenges.

Our focus in this paper is in one of them: the problem of the aggregation of several (imprecise) probability models assigned by different experts into a global one that adequately represents the opinion of the group. This is a problem that has been separately analysed by several authors in terms of different particular cases of imprecise probabilities: coherent sets of desirable gambles, credal sets or coherent lower previsions [7], coherent lower probabilities [1,5,6], probability intervals [5] or possibility measures [3,4]. Because of this, quite a number of different rules as well as several rationality criteria have been considered, often referring to the same notion with different names.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 421–428, 2024.
https://doi.org/10.1007/978-3-031-65993-5_52

Our goal with this paper is to help clarifying and completing the picture drawn by the work in [1,3–7]. Specifically: (i) we look at the problem from the point of view of coherent lower previsions, that include as particular cases the majority of imprecise probability models; (ii) we formulate in this framework twelve of the rationality criteria considered in the studies above; and (iii) analyse five aggregation rules proposed in the literature: the conjunction, disjunction, Pareto, mixture and conjunction-disjunction rules, determining which properties are satisfied by each of them.

The remainder of this contribution is organised as follows: after reviewing the basics of coherent lower previsions in Sect. 2, in Sects. 3 and 4 we introduce the rationality criteria and aggregation rules to be analysed in the paper. Section 5 analyses the properties satisfied by each rule. We conclude the paper in Sect. 6 with some discussion. Due to the space limitations, some proofs and mathematical details have been omitted.

2 Basics of Coherent Lower Previsions

Let us briefly introduce the main notions related to coherent lower prevision we shall consider in this contribution; we refer to [8] for a more detailed account.

Let $\mathcal{X}$ be a possibility space. A *gamble* is a bounded function $f : \mathcal{X} \to \mathbb{R}$. It represents an uncertain reward, meaning that we obtain $f(x)$ if the outcome of an experiment taking values in $\mathcal{X}$ turns out to be x. An instance of gambles are the indicator functions I_A of events $A \subseteq \mathcal{X}$, that take the value 1 on the elements of A and 0 elsewhere. We denote by $\mathcal{L}(\mathcal{X})$ the set of all gambles on $\mathcal{X}$

Given a set of gambles $\mathcal{K} \subseteq \mathcal{L}(\mathcal{X})$, a *lower prevision* is a function $\underline{P} : \mathcal{K} \to \mathbb{R}$. One particular case of interest is when the domain $\mathcal{K}$ coincides with the family of indicator functions $\{I_A : A \subseteq \mathcal{X}\}$. In this case, the lower prevision is called *lower probability* and we write $\underline{P}(A)$ instead of $\underline{P}(I_A)$ for simplicity.

The lower prevision of a gamble f may be understood as a subject's supremum acceptable buying price for this gamble, in the sense that she considers the transaction $f - \mu$ to be desirable for any real-value $\mu < \underline{P}(f)$. Following this interpretation, a lower prevision is called *coherent* when a combination of acceptable transactions does not lead to a sure loss, no matter the outcome of the experiment, and moreover when it does not entail a raise of the supremum acceptable buying price that was previously established for another gamble. When the domain $\mathcal{K}$ is a linear space, coherence holds if and only if:

COH1. $\underline{P}(f) \geq \inf f$ for any $f \in \mathcal{K}$.
COH2. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for any $f \in \mathcal{K}$ and $\lambda > 0$.
COH3. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ for any $f, g \in \mathcal{K}$.

A coherent lower prevision defined on $\mathcal{K}$ can always be extended to $\mathcal{L}(\mathcal{X})$ applying a procedure called *natural extension*, giving rise to the coherent lower prevision

$$\underline{E}(f) = \sup_{g_i \in \mathcal{K}, \lambda_i \geq 0, n \in \mathbb{N}} \inf_{x \in \mathcal{X}} \left\{ f(x) - \sum_{i=1}^{n} \lambda_i \big[g_i(x) - \underline{P}(g_i) \big] \right\} \quad \forall f \in \mathcal{L}(\mathcal{X}). \quad (1)$$

For this reason, we may assume without loss of generality that a coherent lower prevision is defined on $\mathcal{L}(\mathcal{X})$.

A lower prevision $\underline{P}$ always induces an *upper* prevision $\overline{P}$ by conjugacy: $\overline{P}(f) = -\underline{P}(-f)$ whenever $-f \in \mathcal{K}$. The upper prevision of a gamble f represents a subject's infimum acceptable *selling* price for f, in the sense that the transaction $\mu - f$ is considered acceptable for any real value $\mu > \overline{P}(f)$.

When a coherent lower prevision $\underline{P}$ on $\mathcal{L}(\mathcal{X})$ satisfies $\underline{P}(f) = \overline{P}(f)$ for every gamble f, it is called a *linear* prevision, and it corresponds to the expectation operator with respect to its restriction to events, that is a finitely additive probability. Any coherent lower prevision $\underline{P}$ has an associated credal set

$$\mathcal{M}(\underline{P}) := \left\{P \text{ linear prevision} : P(f) \geq \underline{P}(f)\ \forall f \in \mathcal{L}(\mathcal{X})\right\}. \tag{2}$$

Conversely, a lower prevision $\underline{P}$ is coherent if and only if it is the lower envelope of the set $\mathcal{M}(\underline{P})$ it determines via Eq. (2):

$$\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f) \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

A lower prevision $\underline{P}$ *avoids sure loss* when its credal set $\mathcal{M}(\underline{P})$ is non-empty. This is equivalent to the existence of some coherent lower prevision $\underline{Q}$ such that $\underline{Q} \geq \underline{P}$. The smallest such prevision is the natural extension given by Eq. (1).

3 Aggregation of Coherent Lower Previsions

Assume we have a group of n experts, and that each of them models her uncertainty about an experiment in terms of a coherent lower prevision on $\mathcal{L}(\mathcal{X})$. The goal of an aggregation procedure is to determine another lower prevision on $\mathcal{L}(\mathcal{X})$ that adequately summarises the opinions of the group. If we denote by $\underline{\mathbb{P}}(\mathcal{X})$ the set of lower previsions on $\mathcal{X}$, and by $\underline{\mathbb{P}}'(\mathcal{X})$ the subset of coherent lower previsions, we formally define an aggregation rule as follows:

Definition 1. *An* aggregation rule on coherent lower previsions *is a map* $A : \left(\underline{\mathbb{P}}'(\mathcal{X})\right)^n \to \underline{\mathbb{P}}(\mathcal{X})$.

To ease the notation, we shall denote $\underline{P} := A(\underline{P}_1, \ldots, \underline{P}_n)$. A number of desirable properties have been proposed in the literature in order to ensure the rationality of an aggregation rule. The ones analysed in this paper are the following:

Coherence: [1,7] The aggregated model should be a coherent lower prevision. This property, referred to as *convexity* in [3], guarantees that the aggregated model satisfies the same rationality criteria than the original assessments.

Unanimity: [1,5,7] If μ is a desirable buying price for a gamble f for all the experts, it should also be a desirable buying price in the aggregated model. Formally:

$$\underline{P}(f) \geq \min_{i=1,\ldots,n} \underline{P}_i(f) \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

This property also appears in [4] under the name *possibility preservation*.

Reconciliation: [1,7] The supremum acceptable buying price for a gamble should not exceed its infimum acceptable selling price for some expert:

$$\underline{P}(f) \leq \min_{i=1,\ldots,n} \overline{P}_i(f) \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

This property appears in [4] under the name *fairness* and in [3] as *information relevance*.

Indeterminacy: [1,7] The aggregated model should be bounded above by the natural extension of $\max_{i=1,\ldots,n} \underline{P}_i$, which is given by:

$$\sup \inf_{x \in \mathcal{X}} \left\{ f(x) - \sum_{i=1}^{n} f_i(x) : \underline{P}_i(f_i) \geq 0 \ \forall i = 1, \ldots, n \right\} \quad \forall f \in \mathcal{L}(\mathcal{X}). \tag{3}$$

Strong indeterminacy: [1,7] If μ is an acceptable buying price for a gamble in the aggregated model, it should also be acceptable for at least one expert: $\underline{P}(f) \leq \max_{i=1,\ldots,n} \underline{P}_i(f)$ for any $f \in \mathcal{L}(\mathcal{X})$.

Strong Pareto: [7] If μ is an acceptable buying price for at least one expert and it is not an acceptable selling price for any of the experts, then it should also be an acceptable buying price in the aggregated model. Formally:

$$\underline{P}(f) \geq \min \left\{ \max_{i=1,\ldots,n} \underline{P}_i(f), \min_{i=1,\ldots,n} \overline{P}_i(f) \right\} \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

This property is somehow controversial since, as pointed out in [7], it is in general incompatible with coherence.

Conjunction: [1,7] Whenever $\max_{i=1,\ldots,n} \underline{P}_i$ avoids sure loss, the supremum acceptable buying price for a gamble in the aggregated model should be at least as large as its supremum acceptable buying price for each expert. Formally: $\underline{P}(f) \geq \max_{i=1,\ldots,n} \underline{P}_i(f)$ for any gamble $f \in \mathcal{L}(\mathcal{X})$. This property is called *optimism* in [4].

Total reconciliation: [7] The supremum acceptable buying price for a gamble should not exceed the supremum acceptable buying price assigned by any of the experts: $\underline{P}(f) \leq \min_{i=1,\ldots,n} \underline{P}_i(f)$ for any $f \in \mathcal{L}(\mathcal{X})$.

Symmetry: [1,5,7] The aggregation does not depend of the order of the experts. That is, $A\big(\underline{P}_1, \ldots, \underline{P}_n\big) = A\big(\underline{P}_{\sigma(1)}, \ldots, \underline{P}_{\sigma(n)}\big)$ holds for any permutation σ of $\{1, \ldots, n\}$. This property is called *commutativity* in [3,4].

Idempotence: [3] If all experts agree on a model, that should be the aggregate: $A\big(\underline{P}, \ldots, \underline{P}\big) = \underline{P}$. This is called *unanimity* in [6].

Monotonicity: [4] If we consider the partial order on $\underline{\mathbb{P}}(\mathcal{X})$ given by point-wise dominance, the aggregation rule is monotone increasing in each component.

Zero preservation: [3,6] If $\max_{i=1,\ldots,n} \underline{P}_i(A) = 0$, then $\underline{P}(A) = 0$. Equivalently, if $\min_{i=1,\ldots,n} \overline{P}_i(A) = 1$, then $\overline{P}(A) = 1$

4 Aggregation Rules for Coherent Lower Previsions

Next, we recall the aggregation rules from the literature that we shall analyse in this paper, giving their formulation in the case of coherent lower previsions.

Conjunction (A_C) It corresponds [3,7] to the natural extension of the lower prevision $\max\{\underline{P}_1, \ldots, \underline{P}_n\}$, given by Eq. (3). This rule is characterised by coherence and indeterminacy [7, Thm.5.3].

Disjunction (A_D) It is given [3,7] by $A_D(f) = \min\{\underline{P}_1(f), \ldots, \underline{P}_n(f)\}$ for any gamble $f \in \mathcal{L}(\mathcal{X})$. This rule is characterised by coherence, unanimity and total reconciliation [7, Thm.5.2], and is called *convex pooling* in [6].

Mixture (A_M) Given fixed non-negative $\alpha_1, \ldots, \alpha_n$ satisfying $\alpha_1 + \ldots + \alpha_n = 1$, the mixture aggregation rule gives [3,7]

$$A_M(f) = \sum_{j=1}^{n} \alpha_j \underline{P}_j(f) \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

This rule is called *linear opinion pooling* in [6]. As particular cases, it includes the situation where all experts have the same weight in the aggregation ($\alpha_1 = \ldots = \alpha_n = \frac{1}{n}$) as well as the *dictatorship* [2], that corresponds to $\alpha_i = 1$ for one i, and $\alpha_j = 0$ for any $j \neq i$.

Pareto (A_P) The Pareto rule [7] considers a gamble desirable when it is desirable for at least one member of the group and it is not undesirable for any other member. It leads to the following lower prevision:

$$A_P(f) = \min\left\{\max_{j=1,\ldots,n} \underline{P}_j(f), \min_{j=1,\ldots,n} \overline{P}_j(f)\right\} \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

The Pareto rule is characterised by reconciliation, strong indeterminacy and total reconciliation.

Conjunction-Disjunction (A_{CD}) This rule applies the conjunction rule when it leads to a coherent lower prevision (that is, when the lower prevision $\max_{i=1}^{n} \underline{P}_i$ avoids sure loss) and the disjunction rule otherwise [4].

It is not difficult to see that these rules satisfy:

$$A_D \leq A_M, A_P, A_{CD} \leq A_C. \tag{4}$$

To see that there is no dominance relation between A_M, A_P, A_{CD}, note that A_{CD} will either coincide with A_C (if it is coherent) or with A_D, and considering these two cases we can build examples where it is smaller or greater than A_M, A_P. With respect to these two, if it were $A_P \leq A_M$ for all mixtures it should also be $A_P \leq \underline{P}_i$ for any $i = 1, \ldots, n$, meaning that $A_P = A_D$. If it were $A_P \geq A_M$ for all mixtures, in particular $A_P \geq \underline{P}_i$ for all i whence $A_P \geq \max\{\underline{P}_1, \ldots, \underline{P}_n\}$. None of these conditions hold in general, as following example shows:

Example 1. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and consider $\underline{P}_1, \underline{P}_2$ be the coherent lower previsions that are the natural extension of their restriction to events,given by:
The Pareto rule (A_P) does not coincide with neither A_C nor A_D, from which

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\mathcal{X}$
$\underline{P}_1(A)$	0.3	0.3	0.3	0.6	0.6	0.7	1
$\overline{P}_1(A)$	0.3	0.4	0.4	0.7	0.7	0.7	1
$\underline{P}_2(A)$	0.35	0.2	0.2	0.65	0.65	0.5	1
$\overline{P}_2(A)$	0.5	0.35	0.35	0.8	0.8	0.65	1
$\max\{\underline{P}_1(A), \underline{P}_2(A)\}$	0.35	0.3	0.3	0.65	0.65	0.7	1
$\min\{\overline{P}_1(A), \overline{P}_2(A)\}$	0.3	0.35	0.35	0.7	0.7	0.65	1
$A_P(A)$	0.3	0.3	0.3	0.65	0.65	0.65	1
$A_D(A) = A_{CD}(A)$	0.3	0.2	0.2	0.6	0.6	0.5	1

it follows that it does not have a dominance relation with A_M. Also, A_C is not coherent since $\max\{\underline{P}_1, \underline{P}_2\}(\{x_1\}) = 0.35$ and $\max\{\underline{P}_1, \underline{P}_2\}(\{x_2, x_3\}) = 0.7$, which cannot be satisfied by any linear prevision. ♦

5 Rationality Criteria Satisfied by the Rules

Let us discuss which of the rationality criteria are satisfied by the aggregation rules from the previous section. Note that from Eq. (4), any criterion based on inequalities that is satisfied by the conjunction and the disjunction will also be satisfied by the other three rules.

First of all, coherence is preserved under lower envelopes or mixtures. The conjunction rule produces a coherent lower prevision if and only if the lower prevision $\max_{i=1}^n \underline{P}_i$ avoids sure loss, which in turn holds if and only if $\cap_{i=1}^n \mathcal{M}(\underline{P}_i)$ is non-empty. This also tells us that the CD rule always gives a coherent lower prevision, since it is a modification of the conjunction rule to ensure it. As mentioned in [7, Sect. 7.2], the Pareto rule may fail to preserve coherence.

With respect to unanimity, it suffices to note that it is satisfied by the least informative rule, that is disjunction, and as a consequence by all the others.

Reconciliation is satisfied by A_D and A_P. From Eq. (4), A_C satisfies it as long as $\cap_{i=1}^n \mathcal{M}(\underline{P}_i) \neq \emptyset$; Example 1 shows that it will not be the case in general. From these comments it follows that A_{CD} also satisfies reconciliation. Finally, A_M does not satisfy it: considering a dictator ($\alpha_i = 1$), we deduce that it should be $\max_{i=1,\ldots,n} \underline{P}_i \leq \min_{i=1,\ldots,n} \overline{P}_i$, which does not generally hold by Example 1.

Indeterminacy is equivalent to $A \leq A_C$ that, as we have seen in Eq. (4), is true for all the aggregation criteria.

With respect to strong indeterminacy, it is satisfied by A_D, A_M, A_P; in the case of A_C, A_{CD}, it will not hold when $\max_{i=1}^n \underline{P}_i$ avoids sure loss but is not coherent. To see that such a scenario is possible, consider the following example:

Example 2. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and consider the coherent lower previsions $\underline{P}_1$ and $\underline{P}_2$ that are the natural extension of their restriction to events, given by:

A	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\mathcal{X}$
$\underline{P}_1$	0.24	0.24	0.32	0.48	0.56	0.56	1
$\underline{P}_2$	0.32	0.32	0.16	0.64	0.56	0.56	1
$\max\{\underline{P}_1, \underline{P}_2\}$	0.32	0.32	0.32	0.64	0.56	0.56	1
$A_C(\underline{P}_1, \underline{P}_2)$	0.32	0.32	0.32	0.64	0.64	0.64	1

The table above also shows the values of $\max\{\underline{P}_1, \underline{P}_2\}$ and its natural extension, $A_C(\underline{P}_1, \underline{P}_2)$, from which the inequality between A_C and $\max\{\underline{P}_1, \underline{P}_2\}$ ensues. ♦

Strong Pareto means that $A \geq A_P$ and as a consequence, it is satisfied by A_P and by A_C, but not necessarily by A_D, and as a consequence by A_{CD}; an explicit counterexample is Example 1. Since we have seen in Sect. 4 that $A_M \not\geq A_P$ in general, we deduce that strong Pareto is not satisfied by A_M either.

Conjunction means that if A_C is coherent, then $A \geq \max_{i=1}^n \underline{P}_i$. This is satisfied by A_C, A_P and A_{CD}. To see that it is not satisfied by A_D, A_M, it suffices to consider an example where $\min_{i=1}^n \underline{P}_i \neq \max_{i=1}^n \underline{P}_i$ and the latter avoids sure loss. One such instance would be obtained if $\underline{P}_1 \lneq \underline{P}_2$.

For symmetry, we easily see that it is satisfied by A_C, A_D, A_P and A_{CD}; however, it will only be satisfied by A_M when $\alpha_1 = \cdots = \alpha_n$.

Regarding zero preservation, it is satisfied by A_D, A_M and A_P. However, it need not be satisfied by A_C nor A_{CD}. For a counterexample, on $\mathcal{X} = \{x_1, x_2, x_3\}$ consider $\underline{P}_1$ and $\underline{P}_2$ to be the lower envelopes of $\{(0, 0.3, 0.7), (0.2, 0.4, 0.4)\}$ and $\{(0, 0.7, 0.3), (0.2, 0.4, 0.4)\}$. The violation appears on $A = \{x_1\}$.

Finally, idempotence and monotonicity are trivially satisfied by all the rules, while total reconciliation is only satisfied by A_D.

Table 1 summarises our results.

6 Conclusions

In this contribution we have analysed five aggregation rules for aggregating coherent lower previsions and conducted a comparative study by analysing some of the rationality criteria they may satisfy. Among these rules, Pareto stands out as the only rule that does not preserve coherence, making its practical use doubtful. The mixture seems reasonable if we do not want to become too precise/imprecise, and also if we would like to assign weights to the experts. Otherwise, both conjunction and disjunction seem sensible, as well as the conjunction-disjunction rule that circumvents the lack of coherence of the conjunction when the experts are inconsistent.

This contribution presents only the initial steps in our study. As further research, we should analyse not only the twelve properties outlined in Sect. 3 but also additional criteria that can be found in the literature, along with exploring the connections between them. Furthermore, we intend to investigate other

Table 1. Summary of the properties satisfied by each aggregation rule.

Property	Conjunction	Disjunction	Mixture	Pareto	CD
Coherence	✗*	✓	✓	✗	✓
Unanimity	✓	✓	✓	✓	✓
Reconciliation	✗*	✓	✗	✓	✓
Indeterminacy	✓	✓	✓	✓	✓
Strong indeterminacy	✗	✓	✓	✓	✗
Strong Pareto	✓	✗	✗	✓	✗
Conjunction	✓	✗	✗	✓	✓
Total reconciliation	✗	✓	✗	✗	✗
Symmetry	✓	✓	✗	✓	✓
Monotonicity	✓	✓	✓	✓	✓
Idempotence	✓	✓	✓	✓	✓
Zero preservation	✗	✓	✓	✓	✗

(*) It is satisfied when $\max_{i=1,\ldots,n} \underline{P}_i$ avoids sure loss.

aggregation rules such as the maximal consistent sets (MSC) rule [7] or oligarchy [2], as well as investigating how to handle the partial and global conflicts between the sources. This should help the practitioner to have a clearer view when approaching the problem of aggregation under imprecision.

Acknowledgements. This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "FEDER/UE".

References

1. Bradley, S.: Aggregating belief models. Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications, PMLR **103**, 38–48 (2019)
2. Casanova, A., Miranda, E., Zaffalon, M.: Joint desirability foundations of social choice and opinion pooling. Ann. Math. Artif. Intell. **89**, 965–1011 (2021)
3. Destercke, S., Dubois, D., Chojnacki, E.: Possibilistic information fusion using maximal coherent subsets. IEEE Trans. Fuzzy Syst. **17**(1), 79–92 (2009)
4. Dubois, D., Liu, W., Ma, J., Prade, H.: The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. Inform. Fusion 32, 12–39 (2016)
5. Moral, S., del Sagrado, J.: Aggregation of imprecise probabilities. In: Bouchon-Meunier, B. (Ed.) Aggregation and Fusion of Imperfect Information. Studies in Fuzziness and Soft Computing Vol. 12, pp. 162–188. Springer (1998)
6. Stewart, R., Ojea-Quintana, I.: Probabilistic opinion pooling with imprecise probabilities. J. Philos. Log. **47**(1), 17–45 (2018)
7. Walley, P.: The elicitation and aggregation of beliefs. Statistics Research Report 23. Technical report, University of Warwick (1982)
8. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Copulas and Triangular Norms: Selected Commonalities and Differences

Susanne Saminger-Platz(✉)

Institute for Mathematical Methods in Medicine and Data-Based Modeling, Johannes Kepler University Linz, 4020 Linz, Austria
susanne.saminger-platz@jku.at

Abstract. Copulas and triangular norms share a common history in the context of probabilistic metric spaces, but both have roots way back in different mathematical disciplines and continued to develop further in different fields of applications as well as in theory. In this contribution, we discuss some questions posed independently in each respective fields and highlight related results providing a small tour on selected commonalities and differences of copulas and t-norms.

Keywords: Copulas · Triangular norms · Probabilistic metric spaces

1 Introduction

Copulas and triangular norms share a common history as relevant operations in the field of probabilistic metric spaces. Their roots relate back to different mathematical disciplines such as probability theory and statistics, algebra, but also functional equations. Their developments throughout the last decades has led them to seemingly live separate lifes in different fields of mathematics, statistics and logics as well as in different fields of applications like e.g., finance, hydrology resp. preference modeling and decision making, generalized measures and integrals.[1] Nevertheless both classes of functions are quite related to each other anticipating that results from either field will nourish advances in the respective other field. In this contribution, we provide the basic concepts and definitions, a short account on their historic developments and forerunners, explain their common roots and highlight some mathematically interesting results. Since both fields have advanced greatly in the past, this tour on selected commonalities and differences needs to be incomplete and will reflect the author's personal views and experiences with both types of operations.

This abstract provides the basic definitions of both operations and sketches their historic developments along with the main references and monographs for

[1] In order to keep the list of references of this contribution at some reasonable length we have refrained from providing single references at this place but kindly ask to follow the list of references provided in the recommended monographs.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 429–439, 2024.
https://doi.org/10.1007/978-3-031-65993-5_53

basic and further reading. The aim is to allow experts from either of the fields to get attracted by the other field. A glimpse on the topics to be presented at the conference is provided in the last section.

2 Preliminaries on Copulas

The name *copula* for functions linking an n-dimensional distribution with its one-dimensional marginals goes back to 1959 when Abe Sklar [104] proved (for the case $n = 2$) a result which is now referred to as Sklar's Theorem. Links between multivariate distributions and their one-dimensional marginals have been studied before, e.g., by Hoeffding [57,58], Fréchet [44], Dall'Aglio [23–25], and Féron [40], and also later on without any reference to the concept of copulas (see, e.g., [96,107]). In his article *Random Variables, Distribution Functions and Copulas — A Personal Look Backward and Forward* [105] from 1993, A. Sklar reports on and settles his findings into the context of probability and number theory as discussed *up in the early (19)60ties* and illustrates that the important role of copulas in the study of concepts and measures of dependence had been clear from their very early beginnings — besides their importance in the study of probabilistic metric space.

The book *Probabilistic Metric Spaces* [102] published in 1986 by B. Schweizer and A. Sklar, republished with extend results in [103], holds an own chapter on copulas and discusses properties of n-increasing functions, (sub)copulas and joint distribution functions, copulas and t-norms, copulas and random variables as well as margins of copulas. As reported by B. Schweizer in [97] *by the end of the 1980s, the volume of work on copulas and related matters had grown to such an extent that G. Dall'Aglio organized a symposium on distributions with given marginals, which was held in Rome in 1990.*

Several symposia dedicated to *Distributions with given/fixed marginals (and statistical modeling)* followed, up to a meeting in Lecce, Italy, in 2009 celebrating the first 50 years of copulas. As outlined in the preface of [35] *a series of conferences that greatly helped the development of the field, since each of them offered the chance of presenting one's results and to learn those of other researchers.* The findings of the meetings and results discussed have been published in books and conference proceedings ([6,22,26,48,49,91]). Several interviews by and with distinguished colleagues published in *Dependence Modeling* allow the interested reader to obtain personal views on the early advancements of copula theory and the discussion of probabilistic, earlier called statistical metric, spaces (see, e.g., [32–34,37,50–52]).

The first monographs mainly dedicated to copulas got published in the late 1990ties — H. Joe's book *Multivariate models and dependence concepts* [62] and R. Nelsen's *Introduction to copulas* [81,82]. We shall also recommend the more recent and comprehensive monographs [5,36,63] for more details on copulas and their applications, in particular for further references to the various fields of application of copulas. Concluding this part, we briefly summarize the very basics on copulas.

Definition 1. *A function* $C : [0,1]^d \to [0,1]$ *is a* d-copula *if, and only if, the following conditions hold:*

(i) $C(u_1, \ldots, u_d) = 0$ *whenever* $u_j = 0$ *for at least one index* $j \in \{1, \ldots, d\}$,
(ii) when all the arguments of C *are equal to* 1, *but possibly for the* j*-th one, then*

$$C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j,$$

(iii) C *is* d*-increasing i.e., for all* $[\mathbf{a}, \mathbf{b}] \subseteq [0,1]^d$, *its* C*-volume* $V_C([\mathbf{a}, \mathbf{b}])$ *is non-negative, i.e., if*

$$V_C([\mathbf{a}, \mathbf{b}]) := \sum_{\mathbf{v} \in \mathrm{ver}([\mathbf{a}, \mathbf{b}])} \mathrm{sgn}(\mathbf{v}) C(\mathbf{v}) \geq 0$$

where $\mathrm{sgn}(\mathbf{v}) = (-1)^{\mathrm{card}(\{j \in [d] \colon v_j = a_j\})}$ *and* $\mathrm{ver}([\mathbf{a}, \mathbf{b}])$ *denotes the set of vertices of* $[\mathbf{a}, \mathbf{b}]$.

For the bivariate case, it is immediately seen that a copula $C \colon [0,1]^2 \to [0,1]$ satisfies the two boundary conditions $C(u,0) = C(0,u) = 0$ and $C(u,1) = C(1,u) = u$ for all $u \in [0,1]$ and that its 2-increasingness corresponds to

$$C(u,v) + C(u^*, v^*) - C(u, v^*) - C(u^*, v) \geq 0.$$

for all $(u,v), (u^*, v^*) \in [0,1]$ with $(u,v) \leq (u^*, v^*)$ (interpreted coordinate-wisely). We will denote the set of all bivariate copulas by $\mathcal{C}_2$.

The *product* copula $\Pi \colon [0,1]^2 \to [0,1]$ and the *Fréchet-Hoeffding upper and lower bounds* $M \colon [0,1]^2 \to [0,1]$ and $W \colon [0,1]^2 \to [0,1]$ given by, respectively,

$$\Pi(x,y) = xy, \qquad M(x,y) = x, y, \qquad W(x,y) = \max x + y - 1, 0.$$

are often referred to as the three *basic* bivariate copulas and for all binary copulas $C \in \mathcal{C}_2$ we have $W \leq C \leq M$.

Howeover, there is an abundance of different copulas. In the monographs [5, 36, 62, 63, 82] one finds plenty of examples of (one- or two-)parametric families of copulas, on the one hand, and several techniques on how to obtain classes of copulas which can, e.g., be constructed and characterized by functions in one variable (e.g., by additive and/or multiplicative generators in the case of Archimedean copulas or isomorphisms), from multivariate distribution functions (like, e.g., for Gaussian copulas), following geometric considerations (e.g., considering patchwork techniques) or techniques from algebra (involving, e.g., ordinal sums) or measure theory (when it comes, e.g., to copulas with prescribed support) on the other hand.[2]

[2] In order to keep the list of references of this contribution at some reasonable length we have refrained from providing single references at this place but kindly ask to follow the list of refences provided in the recommended monographs.

3 Preliniaries on Triangular Norms

The term *triangular norm* appeared for the first time (with slightly different axioms) in 1942 in Karl Menger's paper *Statistical metrics* [76]. The definition of triangular norms (t-norms for short) as they are used today, goes back to B. Schweizer and A. Sklar [98–100] and has first been published in 1958. Algebraically speaking, t-norms are binary operations on $[0,1]$ such that $([0,1],T,\leq)$ is a commutative, totally ordered monoid (compare [59,70]). The study of associative operations as a specific functional equation on the one hand, and the theory of (special topological) semigroups on the other hand provide two rather independent roots for (continuous) t-norms (see [30] for a very short overview on the historic development and extensive list of references therein).

Several construction methods from the theory of semigroups, such as (isomorphic) transformations (e.g., the additive generators mentioned above) and ordinal sums [16–21,66,101], have been successfully applied to construct whole families of t-norms, starting from a few given prototypical examples. For example, three basic t-norms, the minimum $T_{\mathbf{M}}$, the product $T_{\mathbf{P}}$ and the Łukasiewicz t-norm $T_{\mathbf{L}}$, allow us to construct all continuous t-norms using isomorphic transformations and ordinal sums (see, e.g.,[65,67–69,72,89,108–110]).

In the context of functional equations, the associativity equation is the most influential part of the axiomatic definition of a t-norm, and the earliest source in this context seems to be Abel [1] (see also [2,9,10,60], the monographs by J. Aczél [3,4], and the monograph by C. Alsina, M. J. Frank and B. Schweizer [5]). Major achievements were the full characterization of continuous Archimedean t-norms by additive generators in Ling [72] (for the special case of strict t-norms see [100]) and the solution of a famous functional equation by Frank [42].

The primary importance of t-norms, namely in the fields of probabilistic metric spaces, originates from their role in constructing suitable *triangle functions* needed for an analogon of the triangle inequality in such spaces. At the time of their introduction, it was not yet clear that (continuous resp. left-continuous) t-norms will play an important role as semantic models of the conjunction in many-valued logics.

First ideas of many-valued logics were studied by J. Łukasiewicz around 1920. He started with a three-valued logic, introducing in particular an implication for it (see also [73–75]). For several decades, advances in different directions in non-classical logics, in particular many-valued logics have been made but were mainly considered as a purely theoretical topic. The introduction of fuzzy sets by L. Zadeh in 1965 in [111] and subsequently of fuzzy logic produced a new impact to the investigation of many-valued logics from a logical and algebraic point of view.

The idea to base the intersection of fuzzy sets resp. the conjunction of statements in fuzzy logics on triangular norms has been discussed at several meetings and conferences, in particular at several *Linz Seminars on Fuzzy Set Theory* since their beginnings in 1979. The idea has finally been established and prepared the grounds for the development of the axiomatics of the concept of *basic*

fuzzy logics by P. Hájek [56], in particular on $[0,1]$-valued logics based on continuous (and later also left-continuous) t-norms. In [55], S. Gottwald provides a personal view on the advancements of t-norms and fuzzy logic as many-valued logic in the course of the Linz Seminar series. The extensive *Handbooks of Mathematical Fuzzy Logic — Volumes 1–3* edited by P. Cintula, C. Noguera and P. Hájek resp. C. Fermüller [13–15] provide an excellent and extensive overview and insight into the field of mathematical fuzzy logics from the logical and algebraic viewpoint along with extensive lists of references. We may also additionally recommend the following monographs and edited volumes for an account of the developments in the fields, further reading and references [11,38,47,53,54,85–88,94,109].

The first monograph solely dedicated to *Triangular norms* got published in 2000 by E. P. Klement, R. Mesiar and E. Pap [65]. Also the book *Logical, Algebraic, Analytical and Probabilistic Aspects of Triangular Norms* published in 2005 and edited by E.P. Klement and R. Mesiar [64] and the monograph *Associative Functions. Triangular norms and Copulas* [5] by C. Alsina, M.J. Frank and B. Schweizer published in 2006 are excellent resources for results and further references on and applications of triangular norms.

Concluding this part, we summarize the very basics on triangular norms.

Definition 2. *A* triangular norm *(briefly* t-norm*) is a binary operation T on $[0,1]$ which is commutative, associative, monotone in both arguments and has 1 as neutral element, i.e., it is a function $T\colon [0,1]^2 \to [0,1]$ such that for all $x, y, z \in [0,1]$:*

(i) $T(x,y) = T(y,x)$,
(ii) $T(x,T(y,z)) = T(T(x,y),z)$,
(iii) $T(x,y) \le T(x,z)$ *whenever* $y \le z$,
(iv) $T(x,1) = x$.

We shall denote by $\mathcal{T}$ the set of a t-norms.

The following are the *four basic t-norms*: the minimum $T_{\mathbf{M}}$, the product $T_{\mathbf{P}}$, th Łukasiewicz t-norm $T_{\mathbf{L}}$, and the drastic product $T_{\mathbf{D}}$ all as functions from $[0,1]^2$ into $[0,1]$ are given by, respectively,

$$T_{\mathbf{M}}(x,y) = \min\{x,y\}, \qquad T_{\mathbf{P}}(x,y) = x \cdot y,$$

$$T_{\mathbf{L}}(x,y) = \max\{x+y-1,0\}, \qquad T_{\mathbf{D}}(x,y) = \begin{cases} 0 & \text{if} (x,y) \in [0,1]^2, \\ \min\{x,y\} & \text{otherwise.} \end{cases}$$

From the monotonicity together with 1 serving as neutral element it immediately follows that 0 is an absorbing element to every t-norm T.

Note that $T_{\mathbf{D}} \le T_{\mathbf{L}} \le T_{\mathbf{P}} \le T_{\mathbf{P}}$ and for all $T \in \mathcal{T}$ we have $T_{\mathbf{D}} \le T \le T_{\mathbf{M}}$. The minimum $T_{\mathbf{M}}$ is the only t-norm where each $x \in [0,1]$ is an idempotent element, whereas the product $T_{\mathbf{P}}$ and the Łukasiewicz t-norm $T_{\mathbf{L}}$ are prototypical examples of two important subclasses of t-norms, namely, of the classes of strict resp. nilpotent t-norms.

4 On Some Commonalities and Differences

As is immediately clear from the definitions, copulas and t-norms enjoy the same boundary conditions; also obvious is that $T_{\mathbf{M}} = M$, $T_{\mathbf{P}} = \Pi$, and $T_{\mathbf{L}} = W$, all of them being copulas, while $T_{\mathbf{D}}$ not being a copula. The 2-increasingness implies the continuity of a bivariate copula such that copulas may only be found among continuous t-norms. On the other hand, t-norms are associative such that in turn t-norms may only be found among associative copulas showing that both sets of functions enjoy a non-empty intersection but neither of them is a subset of the other. The following proposition clarifies their relationship (see also [65,79,80,102]).

Proposition 1. *A copula C is also a t-norm if and only if it is associative. A t-norm T is a copula if and only if it is continuous with Lipschitz constant 1.*

Note that the **associativity** of a copula demands it to be a so-called M-ordinal sum of continuous Archimedean t-norms (copulas) which in turn means that it is fully characterized by a set of (convex) additive generators representing the summand operations. Being an M-ordinal sum of this type enforces also the symmetry resp. exchangeability of the copula.

On the other hand, **associativity** for t-norms opens up the way to natuarlly extend to n-ary operations $T^n\colon [0,1]^n \to [0,1]$, also referred to as *serial iterates* in [103]. The behaviour of $x_T^{(n)} = T(x,\ldots,x)$ for $n \to \infty$ relates to several algebraic properties of the t-norm.

Moreover, it reveals insights into the behaviour of the **diagonal** of a t-norm. Whether or not a t-norm resp. a copula is determined uniquely by its values on the diagonal has led to interesting results in the fields of t-norms and copulas. For t-norms, for example, nice counter examples with fractal structure could be found showing that the continuity of a diagonal does neither imply the left- or right-continuity of a t-norm (see,e.g. [7,65,77,106]); for copulas and a given diagonal function, results on bounds could be revealed which when additionally assuming symmetry are also copulas (see, e.g. [28,29,31,39,43,45,46,71,83,84,90]).

Talking about **continuity**, we will point out that though a full characterization of all continuous t-norms as M-ordinal sum of continuous Archimedean t-norms is well-known (see, e.g., [65,78,102]); the characterization of all **left-continuous t-norms** is still an open problem whose solution would be interesting in particular from the perspective of many-valued logics since left-continuous t-norms allow for **adjoint residual implications** (see, e.g. [12,38,41,61]).

It is obvious that throughout the years an abundance of copulas and t-norms, families thereof and several means and tools for constructing copulas and t-norms have been developed. In our presentation, we can discuss only a selection thereof and will focus on examples and families which provided answers to questions motivated from **functional equation** and inequalities like, e.g., the associative equation or the property of **dominance** (see, e.g. [5,92,93,95,102]) or relate to some other interesting mathematical aspects.

References

1. Abel, N.H.: Untersuchung der Functionen zweier unabhängig veränderlichen Größen x und y, wie $f(x, y)$, welche die Eigenschaft haben, daß $f(z, f(x, y))$ eine symmetrische Function von z, x und y ist. J. Reine Angew. Math. **1**, 11–15 (1826)
2. Aczél, J.: Sur les opérations definies pour des nombres réels. Bull. Soc. Math. France **76**, 59–64 (1949)
3. Aczél, J.: Vorlesungen über Funktionalgleichungen und ihre Anwendungen. Birkhäuser, Basel (1961)
4. Aczél, J.: Lectures on Functional Equations and their Applications. Academic Press, New York (1966)
5. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific, Singapore (2006)
6. Beneš, V., Štěpán, J. (eds.): Distributions with Given Marginals and Moment Problems. Kluwer Academic Publishers, Dordrecht (1997)
7. Bézivin, J.P., Tomás, M.S.: On the determination of strict t-norms on some diagonal segments. Aequationes Math. **45**, 239–245 (1993)
8. Borkowski, L. (ed.): J. Łukasiewicz: Selected Works. Studies in Logic and Foundations of Mathematics. North-Holland, Amsterdam (1970)
9. Brouwer, L.E.J.: Die Theorie der endlichen kontinuierlichen Gruppen, unabhängig von den Axiomen von Lie. Math. Ann. **67**, 246–267 (1909)
10. Cartan, É.: La théorie des groupes finis et continus et l'Analysis Situs. Mém. Sci. Math., Vol. 42. Gauthier-Villars, Paris (1930)
11. Cignoli, R., D'Ottaviano, I.M.L., Mundici, D.: Algebraic Foundations of Many-Valued Reasoning. Kluwer Academic Publishers, Dordrecht (2000)
12. Cignoli, R., Esteva, F., Godo, L., Montagna, F.: On a class of left-continuous t-norms. Fuzzy Sets Syst. **131**, 283–296 (2002)
13. Cintula, P., Fermüller, C., Noguera, C. (eds.): Handbook of mathematical fuzzy logic, Vol. 3. Stud. Log. (Lond.), Vol. 58. College Publications, London (2015)
14. Cintula, P., Hájek, P., Noguera, C. (eds.): Handbook of mathematical fuzzy logic, Vo. 1. Stud. Log. (Lond.), Vol. 37. College Publicationss, London (2011)
15. Cintula, P., Hájek, P., Noguera, C. (eds.): Handbook of mathematical fuzzy logic, Vol. 2. Stud. Log. (Lond.), Vol. 38. College Publicationss, London (2011)
16. Clifford, A.H.: Bands of semigroups. Proc. Amer. Math. Soc. **5**, 499–504 (1954)
17. Clifford, A.H.: Naturally totally ordered commutative semigroups. Amer. J. Math. **76**, 631–646 (1954)
18. Clifford, A.H.: Connected ordered topological semigroups with idempotent endpoints. I. Trans. Amer. Math. Soc. **88**, 80–98 (1958)
19. Clifford, A.H.: Totally ordered commutative semigroups. Bull. Amer. Math. Soc. **64**, 305–316 (1958)
20. Clifford, A.H.: Connected ordered topological semigroups with idempotent endpoints. II. Trans. Amer. Math. Soc. **91**, 193–208 (1959)
21. Climescu, A.C.: Sur l'équation fonctionnelle de l'associativité. Bull. École Polytechn. Iassy **1**, 1–16 (1946)
22. Cuadras, C.M., Fortiana, J., Rodríguez-Lallena, J.A. (eds.): Distributions with Given Marginals and Statistical Modelling. Kluwer Academic Publishers, Dordrecht (2002)
23. Dall'Aglio, G.: Sugli estremi dei momenti delle funzioni di ripartizione doppia. Ann. Sc. Norm. Super. Pisa Cl. Sci. Serie III 10, 35–74 (1956)

24. Dall'Aglio, G.: Sulla compatibilità delle funzioni di ripartizione doppia. Rend. Mat. Appl. Serie V **18**, 385–413 (1959)
25. Dall'Aglio, G.: Les fonctions extrêmes de la classe de Fréchet à 3 dimensions. Publ. Inst. Statist. Univ. Paris **9**, 175–188 (1960)
26. Dall'Aglio, G.: Fréchet classes: the beginnings. In: Dall'Aglio et al. [27], pp. 1–12
27. Dall'Aglio, G., Kotz, S., Salinetti, G. (eds.): Advances in Probability Distributions with Given Marginals. Beyond the Copulas, Mathematics and Its Applications, Vol. 67. Kluwer Academic Publishers, Dordrecht (1991)
28. Durante, F., Fernández-Sánchez, J.: On the classes of copulas and quasi-copulas with a given diagonal section. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **19**, 1–10 (2011)
29. Durante, F., Jaworski, P.: Absolutely continuous copulas with given diagonal sections. Comm. Statist. Theory Methods **37**, 2924–2942 (2008)
30. Durante, F., Klement, E.P., Saminger-Platz, S., Sempi, C.: Ordinal sums: from triangular norms to bi- and multivariate copulas. Fuzzy Sets Syst. **451**, 28–64 (2022)
31. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Copulas with given diagonal sections: novel constructions and applications. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **15**, 397–410 (2007)
32. Durante, F., Puccetti, G., Scherer, M.: A journey from statistics and probability to risk theory. an interview with Ludger Rüschendorf. Depend. Model. **3**(1), 182–195 (2015)
33. Durante, F., Puccetti, G., Scherer, M., Vanduffel, S.: Distributions with given marginals: the beginnings. An interview with Giorgio Dall'Aglio. Depend. Model. **4**(1), 237–250 (2016)
34. Durante, F., Puccetti, G., Scherer, M., Vanduffel, S.: My introduction to copulas. An interview with Roger Nelsen. Depend. Model. **5**(1), 88–98 (2017)
35. Durante, F., Sempi, C.: On the characterization of a class of binary operations on bivariate distribution functions. Publ. Math. Debrecen **69**, 47–63 (2006)
36. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton (2015)
37. Durante, F., Puccetti, G., Scherer, M., Vanduffel, S.: Stat Trek. An interview with Christian Genest. Depend. Model. **4**(1), 109–121 (2016)
38. Esteva, F., Godo, L.: Monoidal t-norm based logic: towards a logic for left-continuous t-norms. Fuzzy Sets Syst. **124**, 271–288 (2001)
39. Fernández-Sánchez, J., Úbeda Flores, M.: Constructions of copulas with given diagonal (and opposite diagonal) sections and some generalizations. Depend. Model. **6**, 139–155 (2018)
40. Féron, R.: Sur les tableaux de corrélation dont les marges sont données. Cas de l'espace à trois dimensions. Publ. Inst. Stat. Univ. Paris **5**, 3–12 (1956)
41. Fodor, J.: Left-continuous t-norms in fuzzy logic: an overview. Acta Polytechnica Hungarica **1**(2), 35–47 (2004)
42. Frank, M.J.: On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math. **19**, 194–226 (1979)
43. Frank, M.J.: Diagonals of copulas and Schröder's equation. Aequationes Math. **51**, 150 (1996)
44. Fréchet, M.: Sur les tableaux de corrélation dont les marges sont données. Ann. Univ. Lyon, Sect. A **9**, 53–77 (1951)
45. Fredricks, G.A., Nelsen, R.B.: Copulas constructed from diagonal sections. In: Beneš and Štěpán [6], pp. 129–136

46. Fredricks, G.A., Nelsen, R.B.: The Bertino family of copulas. In: Cuadras et al. [22], pp. 81–91
47. Galatos, N., Jipsen, P., Kowalski, T., Ono, H.: Residuated lattices: an algebraic glimpse at substructural logics, Studies in Logic and the Foundations of Mathematics, vol. 151. Elsevier B. V, Amsterdam (2007)
48. Genest, C.: Proceedings of the conference on dependence modeling: statistical theory and applications to finance and insurance. Canadian J. Statist. **3**(33) (2005)
49. Genest, C.: Proceedings of the conference on dependence modeling: statistical theory and applications to finance and insurance. Insur. Math. Econom. **1**(37) (2005)
50. Genest, C.: A tribute to Abe Sklar. Depend. Model. **9**(1), 200–224 (2021)
51. Genest, C., Puccetti, G.: A journey beyond the Gaussian world. an interview with Harry Joe. Depend. Model. **6**(1), 288–297 (2018)
52. Genest, G., Scherer, M.: The gentleman copulist. An interview with Carlo Sempi. Depend. Model. **8**(1), 34–44 (2020)
53. Goguen, J.A.: The logic of inexact concepts. Synthese **19**, 325–373 (1968/69)
54. Gottwald, S.: Axiomatizations of t-norm based logics-a survey. Soft. Comput. **4**, 63–67 (2000)
55. Gottwald, S.: Fuzzy logic and the Linz Seminar: Themes and some personal reminiscences. In: Saminger-Platz and Mesiar [94], pp. 1–11
56. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer Academic Publishers, Dordrecht (1998)
57. Hoeffding, W.: Maßstabinvariante Korrelationstheorie. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin 5, 181–233 (1940). (Reprinted as "Scale-invariant correlation theory" in Fisher, N.I. and Sen, P.K. (Eds.), *The Collected Works of Wassily Hoeffding*, pp. 57–107. Springer, New York. 1994)
58. Hoeffding, W.: Maßstabinvariante Korrelationsmaße für diskontinuierliche Verteilungen. Archiv für mathematische Wirtschafts- und Sozialforschung 7, 49–70 (1941), (Reprinted as "Scale-invariant correlation measures for discontinuous distribution" in Fisher, N.I. and Sen, P.K. (Eds.), *The Collected Works of Wassily Hoeffding*, pp. 109–133. Springer, New York. 1994)
59. Hofmann, K.H., Lawson, J.D.: Linearly ordered semigroups: Historic origins and A. H. Clifford's influence. London Math. Soc. Lecture Notes, Vol. 231, pp. 15–39. Cambridge University Press, Cambridge (1996)
60. Hosszú, M.: Some functional equations related with the associativity law. Publ. Math. Debrecen **3**, 205–214 (1954)
61. Jenei, S.: How to construct left-continuous triangular norms-state of the art. Fuzzy Sets Syst. **143**, 27–45 (2004)
62. Joe, H.: Multivariate Models and Dependence Concepts, Monographs on Statistics and Applied Probability, vol. 73. Chapman & Hall, London (1997)
63. Joe, H.: Dependence Modeling with Copulas. CRC Press, Boca Raton (2015)
64. Klement, E.P., Mesiar, R. (eds.): Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms. Elsevier, Amsterdam (2005)
65. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms, Trends in Logic. Studia Logica Library, Vol. 8. Kluwer Academic Publishers, Dordrecht (2000)
66. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms as ordinal sums of semigroups in the sense of A. H. Clifford. Semigroup Forum **65**, 71–82 (2002)
67. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms. Position paper I: basic analytical and algebraic properties. Fuzzy Sets Syst. **143**, 5–26 (2004)

68. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms. Position paper II: general constructions and parameterized families. Fuzzy Sets Syst. **45**, 411–438 (2004)
69. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms. Position paper III: continuous t-norms. Fuzzy Sets Syst. **145**, 439–454 (2004)
70. Klement, E.P., Mesiar, R., Pap, E.: Semigroups and triangular norms. In: Klement and Mesiar [64], chap. 3, pp. 63–93
71. Kokol Bukovšek, D., Mojškerc, B., Stopar, N.: Exact upper bound for copulas with a given diagonal section. Fuzzy Sets Syst. **480**, 108865 (2024)
72. Ling, C.H.: Representation of associative functions. Publ. Math. Debrecen **12**, 189–212 (1965)
73. Łukasiewicz, J.: O logice trówartosciowej. Ruch Filozoficzny **5**, 170–171 (1920), (English translation contained in [8])
74. Łukasiewicz, J.: Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls. Comptes Rendus Séances Société des Sciences et Lettres Varsovie cl. **III**(23), 51–77 (1930)
75. Łukasiewicz, J., Tarski, A.: Untersuchungen über den Aussagenkalkül. Comptes Rendus Séances Société des Sciences et Lettres Varsovie cl. **III**(23), 1–21 (1930)
76. Menger, K.: Statistical metrics. Proc. Nat. Acad. Sci. U.S.A. **8**, 535–537 (1942)
77. Mesiarová-Zemánková, A.: T-norms and t-conorms continuous around diagonals. Fuzzy Sets Syst. **299**, 105–112 (2016)
78. Mostert, P.S., Shields, A.L.: On the structure of semigroups on a compact manifold with boundary. Ann. of Math., II. Ser. **65**, 117–143 (1957)
79. Moynihan, R.: Infinite τ_T products of probability distribution functions. J. Austral. Math. Soc. (Ser. A) **26**, 227–240 (1978)
80. Moynihan, R.: On τ_T semigroups of probability distribution functions II. Aequationes Math. **17**, 19–40 (1978)
81. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics, vol. 139. Springer, New York (1999)
82. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics, vol. 139, 2nd edn. Springer, New York, (2006)
83. Nelsen, R.B., Fredricks, G.A.: Diagonal copulas. In: Beneš and Štěpán [6], pp. 121–127
84. Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M.: On the construction of copulas and quasi-copulas with given diagonal sections. Insurance: Mathematics and Economics 42, 473–483 (2008)
85. Novák, V.: First-order fuzzy logic. Stud. Logica. **46**, 87–109 (1987)
86. Pavelka, J.: On fuzzy logic. I. Many-valued rules of inference. Z. Math. Logik Grundlag. Math. **25**, 45–52 (1979)
87. Pavelka, J.: On fuzzy logic. II. Enriched residuated lattices and semantics of propositional calculi. Z. Math. Logik Grundlag. Math. **25**, 119–134 (1979)
88. Pavelka, J.: On fuzzy logic. III. Semantical completeness of some many-valued propositional calculi. Z. Math. Logik Grundlag. Math. **25**, 447–464 (1979)
89. Petrík, M., Vetterlein, T.: Rees coextensions of finite tomonoids and free pomonoids. Semigroup Forum **99**(2), 345–367 (2019)
90. Quesada-Molina, J.J., Saminger-Platz, S., Sempi, C.: Quasi-copulas with a given sub-diagonal section. Nonlinear Anal. **69**, 4654–4673 (2008)
91. Rüschendorf, L., Schweizer, B., Taylor, M.D. (Eds.): Distributions with Fixed Marginals and Related Topics, Lecture Notes — Monograph Series, Vol. 28. Institute of Mathematical Statistics, Hayward, CA (1996), (proceedings of the AMS-IMS-SIAM Joint Summer Research Conference on Distributions with Fixed

Marginals, Doubly Stochastic Measures and Markov Operators held in Seattle, WA, August 1–5, 1993)
92. Saminger-Platz, S.: The dominance relation in some families of continuous Archimedean t-norms and copulas. Fuzzy Sets Syst. **160**, 2017–2031 (2009)
93. Saminger-Platz, S., De Baets, B., De Meyer, H.: Differential inequality conditions for dominance between continuous Archimedean t-norms. Math. Inequal. Appl. **12**, 191–208 (2009)
94. Saminger-Platz, S., Mesiar, R. (eds.): On Logical, Algebraic, and Probabilistic Aspects of Fuzzy Set Theory, Studies in Fuzziness and Soft Computing, vol. 336. Springer, Switzerland (2016)
95. Sarkoci, P.: Dominance is not transitive on continuous triangular norms. Aequationes Math. **75**, 201–207 (2008)
96. Schwarz, G.: Multivariate distributions with uniformly distributed projections. Ann. Probab. **13**, 1371–1372 (1985)
97. Schweizer, B.: Introduction to copulas. J. Hydrol. Eng. **12**(4) (2007)
98. Schweizer, B., Sklar, A.: Espaces métriques aléatoires. C. R. Acad. Sci. Paris Sér. A **247**, 2092–2094 (1958)
99. Schweizer, B., Sklar, A.: Statistical metric spaces. Pacific J. Math. **10**, 313–334 (1960)
100. Schweizer, B., Sklar, A.: Associative functions and statistical triangle inequalities. Publ. Math. Debrecen **8**, 169–186 (1961)
101. Schweizer, B., Sklar, A.: Associative functions and abstract semigroups. Publ. Math. Debrecen **10**, 69–81 (1963)
102. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. North-Holland, New York (1983)
103. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. Dover Publications, Mineola (2006)
104. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
105. Sklar, A.: Random variables, distribution functions, and copulas—a personal look backward and forward. In: Rüschendorf et al. [91], pp. 1–14, (proceedings of the AMS-IMS-SIAM Joint Summer Research Conference on Distributions with Fixed Marginals, Doubly Stochastic Measures and Markov Operators held in Seattle, WA, August 1–5, 1993)
106. Smutná, D.: Non-continuous t-norms with continuous diagonal. J. Electrical Eng. 51(12/s), 51–53 (2000)
107. Tchen, A.H.: Inequalities for distributions with given marginals. Ann. Probab. **8**, 814–827 (1980)
108. Vetterlein, T., Petrík, M.: The semantics of fuzzy logics: Two approaches to finite tomonoids. In: Saminger-Platz and Mesiar [94], pp. 83–108
109. Vetterlein, T.: The construction of left-continuous t-norms: a geometric approach in two dimensions. Fuzzy Sets Syst. **252**, 1–24 (2014)
110. Vetterlein, T.: Real coextensions as a tool for constructing triangular norms. Inform. Sci. **348**, 357–376 (2016)
111. Zadeh, L.A.: Fuzzy sets. Inform. Control **8**, 338–353 (1965)

Characterizations of Multivariate Implicit Dependence Copulas

Tippawan Santiwipanont, Songkiat Sumetkijakan(✉), and Noppawit Yanpaisan

Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
{tippawan.s,songkiat.s}@chula.ac.th, noppawit.yps@gmail.com

Abstract. The copula C of continuously distributed random variables $X_1, \ldots, X_d$ is said to be an *implicit dependence copula* if there are Borel functions $\alpha_1, \ldots, \alpha_d$ such that $\alpha_1(X_1), \ldots, \alpha_d(X_d)$ are equal almost surely and continuously distributed, that is their common distribution function is continuous. Bivariate implicit dependence copulas have recently been characterized in terms of a generalized Markov product. In this manuscript, the characterizations are extended to the multivariate case in terms of a product of d copulas, called $\mathscr{A}$-product where $\mathscr{A}$ is a class of copulas A_t, $t \in [0,1]$. The class of implicit dependence d-copulas are characterized as $\mathscr{A}$-products of d complete dependence copulas. Explicit forms of the joining copulas A_t are obtained when the functions α_i are countably piecewise monotonic surjections.

Keywords: Implicit dependence copulas · Generalized markov product · Countably piecewise monotonic surjections · Conditional copulas

1 Introduction and Background

By Sklar's theorem, dependence among continuous random variables is inscribed in their unique copula [6,7], while the marginals are considered mere rank-preserving transformations. As such, copulas are simply joint distribution functions of uniform-$(0,1)$ random variables. In the bivariate case, *complete dependence* (CD) copulas are copulas of two continuous random variables that are completely dependent in the sense that one is almost surely a function of the other. According to [2,10,13], the copula of two continuous random variables X and Y is said to be an *implicit dependence* (ID) copula if there exist Borel functions f and g such that $f(X) = g(Y)$ a.s. and $f(X)$ has continuous distribution. Though limited, CD copulas have proved useful in solving theoretical problems [12,15] and have quite simple forms [14]: they are either $C_{e,f}$ or $C_{f,e}$ for some f in $\mathscr{T}$, the class of measure-preserving transformations on $[0,1]$, where e denotes

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 440–448, 2024.
https://doi.org/10.1007/978-3-031-65993-5_54

the identity on $[0,1]$. It is well-known [4,16] that every copula is of the form $C_{f,g}$ for $f, g \in \mathscr{T}$ and that $C_{f,g} = C_{f,e} * C_{e,g}$ where the *Markov product* $*$ makes the class $\mathscr{C}$ of copulas a monoid with zero. See [3,6] for their definitions and properties. A *generalized Markov product* [5,11] of a pair of copulas C, D with respect to a class of *joining copulas* $\mathscr{A} = \{A_t : t \in [0,1]\}$ is the copula defined by $C *_{\mathscr{A}} D(x,y) := \int_0^1 A_t\left(\partial_2 C(x,t), \partial_1 D(t,y)\right) dt$. Recently [8,9], it was proved that every ID copula is a generalized Markov product of some CD copulas $C_{e,f}$ and $C_{g,e}$, in that order, where the joining copulas are conditional joint distributions which, under some simplifying conditions, can be written in a closed form involving the Markov kernel corresponding to the IDC. And the converse is also valid.

In this manuscript, the notion of implicit dependence copulas is extended to a multivariate case in a natural way. They are copulas of $\boldsymbol{\mathcal{U}}(0,1)$ random variables $X_1, \ldots, X_d$ for which there exist $\alpha^{(1)}, \ldots, \alpha^{(d)} \in \mathscr{T}$ such that $\alpha^{(1)}(X_1), \ldots, \alpha^{(d)}(X_d)$ are equal a.s. with continuous distribution function. In the same spirit as the bivariate case, two characterizations of multivariate implicit dependence copulas are obtained. The first approach in Sect. 2 is based on Theorem 2.7 in [1] by Ansari and Rüschendorf which relies on Sklar's theorem and the disintegration theorem. As a result, the joining copulas are conditional joint distributions whose existence is guaranteed by Sklar's theorem. The second approach in Sect. 3 is more constructive as it gives an explicit form of the joining copulas.

2 An Approach Using Sklar's Theorem

This section gives a full characterization of all multivariate implicit dependence copulas. Its proof largely mimics that of Theorem 2.7 in [1] with some modifications. Nonetheless, we include the proof here for completeness.

For random variables $X_1, \ldots, X_d$ and Z, define $F_{(X_1,\ldots,X_d)|Z=t}(x_1, \ldots, x_d) := \mathbb{P}\left((X_1, \ldots, X_d) \in \bigtimes_{i=1}^d (-\infty, x_i] \mid Z = t\right)$, $F_{i|t} := F_{X_i|Z=t}$ for each i and $C_{(X_1,\ldots,X_d)}$ as a copula of $X_1, \ldots, X_d$. Also denote $\mathscr{C}_d$ and $\mathscr{T}$ as the classes of d-copulas and measure-preserving transformations on $\mathbb{I}$, respectively. For every $C_1, \ldots, C_d \in \mathscr{C}_2$ and $\mathcal{A} := \{A_t\}_{t\in\mathbb{I}} \subseteq \mathscr{C}_d$, denote $\underset{i=1}{\overset{d}{\circledast}}_{\mathcal{A}} C_i(x_1, \ldots, x_d) := \int_0^1 A_t\left((\partial_2 C_i(x_i,t))_{1\le i\le d}\right) \mathrm{d}t$ whenever the integrand is measurable in t. Notice that for each $C \in \mathscr{C}_2$ and $t \in \mathbb{I}$, $\partial_2 C(x,t)$ can be extended to a distribution function $F_t(x) := \lim_{w \searrow x} \partial_2 C(w,t)$. In fact, $F_t(x) = K_{C^T}(t, [0,x])$ where K_C denotes the *Markov kernel* of C [6] and C^T denotes the *transpose* of C. Moreover, define $G^{(-1)}(x) := \inf\{u : G(u) \ge x\}$.

Theorem 1. *Let* $\alpha^{(1)}, \ldots, \alpha^{(d)} \in \mathscr{T}$.

1. *If* $\mathcal{A} = \{A_t\}_{t\in\mathbb{I}}$ *is a class of* d*-copulas such that the map* $t \mapsto A_t(x_1, \ldots, x_d)$ *is Borel measurable for all* $(x_1, \ldots, x_d) \in \mathbb{I}^d$*, then* $C := \underset{i=1}{\overset{d}{\circledast}}_{\mathcal{A}} C_{e,\alpha^{(i)}}$ *is an IDC. In fact,* $C = C_{(X_1,\ldots,X_d)}$ *for some* $X_1, \ldots, X_d \sim \boldsymbol{\mathcal{U}}(0,1)$ *satisfying* $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ *a.s.*

2. *Conversely, if* C *is the copula of* $X_1, \ldots, X_d \sim \boldsymbol{\mathcal{U}}(0,1)$ *and* $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ *a.s., then there exists* $\mathcal{A} = \{A_t\}_{t\in\mathbb{I}} \subseteq \mathscr{C}_d$ *such that* $C = \overset{d}{\underset{i=1}{\circledast_{\mathcal{A}}}} C_{e,\alpha^{(i)}}$. *More precisely, for* $Z \sim \boldsymbol{\mathcal{U}}(0,1)$ *with* $C_{(X_1,Z)} = C_{e,\alpha^{(1)}}$, *we have* $C_{(X_i,Z)} = C_{e,\alpha^{(i)}}$ *for all* i *and* A_t *defined on* $\mathbb{I}^2$ *for a.e.* $t \in \mathbb{I}$ *by*

$$A_t(x_1, \ldots, x_d) := F_{(X_1,\ldots,X_d)|Z=t}(F^{(-1)}_{1|t}(x_1), \ldots, F^{(-1)}_{d|t}(x_d)) \tag{1}$$

is a possible choice of joining copula, where $F_{i|t}(x_i) = K_{\alpha^{(i)}}(t, [0, x_i]) := K_{C_{\alpha^{(i)},e}}(t, [0, x_i])$, *for* $i = 1, \ldots, d$.

Proof. 1. It can be divided into 2 subparts.

(1.1) Since $K_{\alpha^{(i)}}(t, [0, x_i])$ is measurable in t, $A_t\left(\left(\partial_2 C_{e,\alpha^{(i)}}(x_i, t)\right)_{i=1}^{d}\right)$ is measurable in t. Consequently, C is well-defined and hence a copula. Let μ_C be the Borel probability measure extension of V_C to $\mathscr{B}(\mathbb{I}^d)$. For each $i = 1, \ldots, d$, define the random variable X_i as the projection map of $\mathbb{I}^d$ onto the i^{th}-coordinate. Since μ_C has uniform marginals, each X_i is uniformly distributed on $[0, 1]$. Also, their joint distribution function is C, that is, $C_{(X_1,\ldots,X_d)} = C$.

(1.2) Evidently, $D(x_1, \ldots, x_d, z) := \int_0^z A_t\left(F_{1|t}(x_1), \ldots, F_{d|t}(x_d)\right) \mathrm{d}t$ is a $(d+1)$-copula whose marginal on the first d components is $C = \overset{d}{\underset{i=1}{\circledast_{\mathcal{A}}}} C_{e,\alpha^{(i)}} = C_{(X_1,\ldots,X_d)}$. Hence, D is the copula of $X_1, \ldots, X_d, Z$ for some $Z \sim \boldsymbol{\mathcal{U}}(0,1)$. Moreover, the $(i, d+1)$-marginal is $C_{X_i,Z} = C_{e,\alpha^{(i)}}$, i.e. $Z = \alpha^{(i)}(X_i)$ a.s., for all i. So C is an IDC with relation $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ a.s. as desired.

2. From $C_{(X_1,Z)} = C_{e,\alpha^{(1)}}$, we have $Z = \alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ a.s. and by considering the relation $Z = \alpha^{(i)}(X_i)$ a.s., we have $C_{(X_i,Z)} = C_{e,\alpha^{(i)}}$ for all i. Next, from (1) in which the existence of A_t can also be proved by Sklar's theorem, $F_{(X_1,\ldots,X_d)|Z=t}(x_1, \ldots, x_d) = A_t\left(\left(F_{i|t}(x_i)\right)_{1\le i\le d}\right)$ which implies that

$$C(x_1, \ldots, x_d) = \int_0^1 A_t\left(\left(\partial_2 C_{e,\alpha^{(i)}}(x_i, t)\right)_{1\le i\le d}\right) \mathrm{d}t = \left(\overset{d}{\underset{i=1}{\circledast_{\mathcal{A}}}} C_{e,\alpha^{(i)}}\right)(x_1, \ldots, x_d).$$

3 A Constructive Method

The main theorem of this section gives a partial characterization of multivariate implicit dependence copulas. The measure-preserving transformations $\alpha^{(1)}, \ldots, \alpha^{(d)}$ need to be countably piecewise monotonic surjections (CPMS) and the proof follows the same line as those of Thms. 5 and 10 in [9]. Even though the following theorem is stated for d-copulas, it shall be proved for 3-copulas which sufficiently shows how the proof can be extended to arbitrary dimension. We say that a measure-preserving transformation α is a CPMS, denoted by $\alpha \in \mathscr{T}_{\mathrm{P}}$, if there is a partition $P := \{I_n\}_{n\in\mathbb{N}}$ consisting of open intervals such that each $\alpha_n := \alpha|_{I_n}$ is a strictly monotonic function from I_n onto $(0, 1)$. Furthermore, a d-copula C is said to be in $\mathscr{C}_{\mathrm{PID}}$ if there are $X_1, \ldots, X_d \sim \boldsymbol{\mathcal{U}}(0,1)$ and $\alpha^{(1)}, \ldots, \alpha^{(d)} \in \mathscr{T}_{\mathrm{P}}$ satisfying $C = C_{(X_1,\ldots,X_d)}$ and $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ a.s.

Example 1. If $\alpha^{(i)}(x) = \Lambda(x) := 2\min(x, 1-x)$ for $x \in [0,1]$ and $i = 1,2,3$, then every 3-copula C of $X_1, X_2, X_3 \sim \boldsymbol{\mathcal{U}}(0,1)$ for which all $\alpha^{(i)}(X_i)$ are equal a.s. is supported in the union of four diagonals joining $a := (a_1, a_2, a_3)$ and (b_1, b_2, b_3), where $a \in \{(0,0,0), (0,0,1), (0,1,0), (1,0,0)\}$ and $\{a_i, b_i\} = \{0,1\}$ for $i = 1,2,3$.

Remark 1 ([9]). Suppose that $\alpha \in \mathscr{T}_{\mathrm{P}}$ with the partition $P := \{I_n\}_{n\in\mathbb{N}}$.

1. For $m, n \in \mathbb{N}$, define $\prec$ and $\preceq$ on P by $I_m \prec I_n$ if and only if I_m is on the left of I_n and $I_m \preceq I_n$ if and only if $I_m \prec I_n$ or $I_m = I_n$.
2. For a.e. $s \in \mathbb{I}$ and $i \in \mathbb{N}$, define $s_{(i)}$ as the unique number in I_i such that $\alpha(s_{(i)}) = \alpha(s)$. That is, $s = s_{(i)}$ if and only if $s \in I_i$. Consequently, if $s \in I_k$ then $s_{(i)} = \alpha_{ik}(s)$ where $\alpha_{ik} := \alpha_i^{-1} \circ \alpha_k$ ($\alpha_k = \alpha|_{I_k}$). Moreover, for a.e. $u, v \in \mathbb{I}$, if α_i and α_j are either both increasing or both decreasing, then $u_{(i)} < v_{(i)}$ is equivalent to $u_{(j)} < v_{(j)}$; otherwise, $u_{(i)} < v_{(i)}$ is equivalent to $u_{(j)} > v_{(j)}$.
3. For a.e. $x, s \in \mathbb{I}$ with $x \in I_k$, $\partial_2 C_{e,\alpha}(x, \alpha(s)) = \sum_{i=1}^{\infty} \frac{1}{|\alpha'(s_{(i)})|} \mathbb{1}_{(0,x)}(s_{(i)})$. That is, if $\mu_k := \sum_{I_i \preceq I_k} \frac{1}{|\alpha'(s_{(i)})|}$ and $\mu_k^* := \sum_{I_i \prec I_k} \frac{1}{|\alpha'(s_{(i)})|}$ then $\partial_2 C_{e,\alpha}\left(x_{(k)}, \alpha(s)\right) = \mu_k \mathbb{1}_{[0,x_{(k)})}(s_{(k)}) + \mu_k^* \mathbb{1}_{(x_{(k)},1]}(s_{(k)})$. Moreover, since the collection $\mathscr{I}$ of indexed sets $\{i : I_i \prec I_k\}$ and $\{i : I_i \preceq I_k\}$, $k \in \mathbb{N}$, is totally ordered, for every value $\mu \in \{\mu_k, \mu_k^* : k \in \mathbb{N}\}$ there exists $\mathcal{I}_\mu \in \mathscr{I}$ such that $\mu = \sum_{i \in \mathcal{I}_\mu} \frac{1}{|\alpha'(s_{(i)})|}$.

Theorem 2. *Let $\alpha^{(1)}, \ldots, \alpha^{(d)} \in \mathscr{T}_P$ with the corresponding partitions $\left\{I_j^{(i)}\right\}_{j\in\mathbb{N}}$ for each $i = 1, \ldots, d$.*

1. *If $\mathcal{A} := \{A_t\}_{t\in\mathbb{I}} \subset \mathscr{C}_d$ is such that $A_t(x_1, \ldots, x_d)$ is Borel measurable in t for each $(x_1, \ldots, x_d) \in \mathbb{I}^d$, then $C := \underset{i=1}{\overset{d}{\circledast}}_{\mathcal{A}} C_{e,\alpha^{(i)}}$ is an IDC. In fact, C is the copula of some $X_1, \ldots, X_d \sim \boldsymbol{\mathcal{U}}(0,1)$ such that $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ a.s.*
2. *Conversely, if C is the copula of $X_1, \ldots, X_d \sim \boldsymbol{\mathcal{U}}(0,1)$ satisfying $\alpha^{(1)}(X_1) = \cdots = \alpha^{(d)}(X_d)$ a.s., then there exists $\mathcal{A} = \{A_t\}_{t\in\mathbb{I}} \subset \mathscr{C}_d$ such that $C := C_{(X_1,\ldots,X_d)} = \underset{i=1}{\overset{d}{\circledast}}_{\mathcal{A}} C_{e,\alpha^{(i)}}$. More precisely, if we use Remark 1 (2,3) to denote $\alpha := \alpha^{(1)}$, $s := s_1^{(1)}$, $s_i := s_i^{(1)}$, $\mu^{(i)} := \partial_2 C_{e,\alpha^{(i)}}(x^{(i)}, \alpha(s))$ for some $x^{(i)} \in \mathbb{I}$, a choice of d-subcopulas $A_{\alpha(s)}$, for $s \in I_1^{(1)}$, is*

$$A_{\alpha(s)}\left(\left(\mu^{(i)}\right)_{i=1}^{d}\right) = \sum_{j \in I_{\mu^{(1)}}^{(1)}} \frac{1}{|\alpha_j'(s_j)|} K_C\left(s_j, \bigtimes_{k=2}^{d} \bigcup_{\ell \in I_{\mu^{(k)}}^{(k)}} \left[0, \left(\alpha_\ell^{(k)}\right)^{-1} \alpha(s)\right]\right). \tag{2}$$

We prove in the case $d = 3$ with $\alpha, \beta, \gamma \in \mathscr{T}_{\mathrm{P}}$. For the function $\alpha \in \mathscr{T}_{\mathrm{P}}$ with partition $P := \{I_i := (a_i, b_i)\}_{i\in\mathbb{N}}$ of $\mathbb{I}$ on the X-axis, we use $\alpha_i, \mu_i, \mu_i^*, s_{(i)}, I_\mu$ as in Remark 1. In a similar manner, for a function $\beta \in \mathscr{T}_{\mathrm{P}}$ with partition $Q :=$

$\{J_j := (c_j, d_j)\}_{j\in\mathbb{N}}$ of $\mathbb{I}$ on the Y-axis, the notations $\beta_j, \eta_j, \eta_j^*, t^{(j)}, \mathcal{J}_\eta$ are the analogs of $\alpha_i, \mu_i, \mu_i^*, s_{(i)}, \mathcal{I}_\mu$, respectively; for a function $\gamma \in \mathscr{T}_\mathrm{P}$ with partition $R := \{K_k := (e_k, f_k)\}_{k\in\mathbb{N}}$ of $\mathbb{I}$ on the Z-axis, the notations $\gamma_k, \sigma_k, \sigma_k^*, u^{(k)}, \mathcal{K}_\sigma$ are the analogs of $\alpha_i, \mu_i, \mu_i^*, s_{(i)}, \mathcal{I}_\mu$, respectively.

Proof (Proof of Theorem 2 (1)). The fact that C is a 3-copula and $C = C_{(X,Y,Z)}$ for some $X, Y, Z \sim \boldsymbol{\mathcal{U}}(0,1)$ can be shown in the same way as (1) in Theorem 1. So it remains to show that $\alpha \circ X(x,y,z) = \beta \circ Y(x,y,z) = \gamma \circ Z(x,y,z)$ for μ_C-a.e. $(x,y,z) \in \mathbb{I}^3$, which will be shown on each $A_{ijk} := I_i \times J_j \times K_k = (a_i, b_i) \times (c_j, d_j) \times (e_k, f_k)$ that $V_C(A_{ijk} \backslash G) = 0$ where $G := \{\alpha \circ X = \beta \circ Y = \gamma \circ Z\} = \big\{(x,y,z) \in \mathbb{I}^3 \colon \alpha_i(x) = \beta_j(y) = \gamma_k(z) \text{ for some } i,j,k\big\}$. Notice that there are four possibilities of $G_{ijk} := G \cap A_{ijk}$ as illustrated in Fig. 1, depending on whether $\alpha_i, \beta_j, \gamma_k$ are increasing or decreasing.

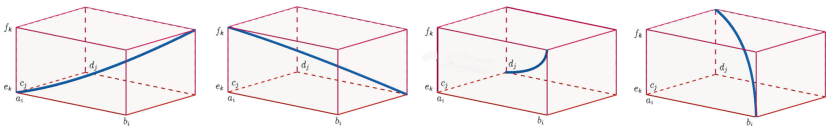


Fig. 1. All possibilities of $G_{ijk} := G \cap A_{ijk}$

Next, if $B = (x_1, x_2) \times (y_1, y_2) \times (z_1, z_2)$ satisfies $\pi_\ell(G_{ijk}) \cap \pi_\ell(B) \neq \emptyset$ for all $\ell \in \{XY, YZ, XZ\}$ (where π_ℓ is the projection on plane ℓ), then by denoting $R_i^\alpha := \big(\min\{\alpha_i(x_1), \alpha_i(x_2)\}, \max\{\alpha_i(x_1), \alpha_i(x_2)\}\big)$, and R_j^β, R_k^γ in the similar manner, we have $R_i^\alpha \cap R_j^\beta$, $R_i^\alpha \cap R_k^\gamma$ and $R_j^\beta \cap R_k^\gamma$ are all nonempty sets. Moreover, since R_i^α, R_j^β and R_k^γ are open intervals, we obtain $R_i^\alpha \cap R_j^\beta \cap R_k^\gamma \neq \emptyset$. That is, any t in the intersection of those three sets satisfies $t = \alpha_i(x) = \beta_j(y) = \gamma_k(z)$ for some $(x,y,z) \in B$ and then $G_{ijk} \cap B \neq \emptyset$. Thus any rectangular box in $A_{ijk} \backslash G$ satisfies $\pi_\ell(G_{ijk}) \cap \pi_\ell(B) = \emptyset$ for some plane $\ell \in \{XY, YZ, XZ\}$.

Furthermore, by the monotonicity of $\pi_\ell(G_{ijk})$, every rectangular box of $A_{ijk} \backslash G$ can be written as a combination of no more than eight boxes in $A_{ijk} \backslash G$ all of which have the common vertex at a corner of A_{ijk}. It then suffices to show that $V_C(B) = 0$ whenever B is such a rectangular box. For a.e. $(x,y,z) \in I_p \times J_q \times K_r$ and a.e. $t \in (0,1)$, the partial derivatives for $C_{e,\alpha}$, $C_{e,\beta}$ and $C_{e,\gamma}$ can be rewritten simply as combinations of indicator functions, of which the coefficients depend on whether α_p, β_q and γ_r are increasing or decreasing as follows:

$$\begin{aligned}
\partial_2 C_{e,\alpha}(x,t) &= \hat{\mu}_p \mathbb{1}_{[0,\alpha(x))}(t) + \check{\mu}_p \mathbb{1}_{(\alpha(x),1]}(t) \quad \text{with } \{\hat{\mu}_p, \check{\mu}_p\} = \{\mu_p, \mu_p^*\} \\
\partial_2 C_{e,\beta}(y,t) &= \hat{\eta}_q \mathbb{1}_{[0,\beta(y))}(t) + \check{\eta}_q \mathbb{1}_{(\beta(y),1]}(t) \quad \text{with } \{\hat{\eta}_q, \check{\eta}_q\} = \{\eta_q, \eta_q^*\} \text{ and,} \\
\partial_2 C_{e,\gamma}(z,t) &= \hat{\sigma}_r \mathbb{1}_{[0,\gamma(z))}(t) + \check{\sigma}_r \mathbb{1}_{(\gamma(x),1]}(t) \quad \text{with } \{\hat{\sigma}_r, \check{\sigma}_r\} = \{\sigma_r, \sigma_r^*\}
\end{aligned}$$

where $\hat{\mu}_p = \mu_p$ if and only if α_p is increasing; while the others are similar. In particular, $\partial_2 C_{e,\alpha}(a_p, t) = \mu_p^*$, $\partial_2 C_{e,\alpha}(b_p, t) = \mu_p$, $\partial_2 C_{e,\beta}(c_q, t) = \eta_q^*$, $\partial_2 C_{e,\beta}(d_q, t) = \eta_q$, $\partial_2 C_{e,\gamma}(e_r, t) = \sigma_r^*$ and $\partial_2 C_{e,\gamma}(f_r, t) = \sigma_r$. The symbols μ_p and μ_p^* are defined in Remark 1 (3) while $\eta_q, \eta_q^*, \sigma_r, \sigma_r^*$ are defined similarly.

From now on, we denote $F_w^{\varphi}(t) := \partial_2 C_{e,\varphi}(w,t)$, where $w \in \{x,y,z\}$ and $\varphi \in \{\alpha,\beta,\gamma\}$, for shortness and suppose for simplicity that α_p, β_q and γ_r are increasing. The other possibilities of α_p, β_q and γ_r can be proved similarly. Moreover, we show only the case $\pi_{XY}(G_{pqr}) \cap \pi_{XY}(B) = \emptyset$ since the other cases are similar.

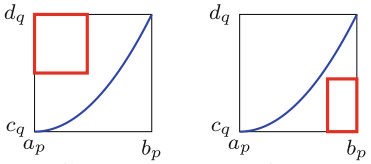


Fig. 2. Possible $\pi_{XY}(G_{pqr})$ and $\pi_{XY}(B)$ under consideration

We shall prove only case 1 where $\pi_{XY}(B) = [a_p, x] \times [y, d_q]$; $\beta_q(y) > \alpha_p(x)$ as case 2 is similar. $V_C(B) \leq V_C(\pi_{XY}(B) \times K_r) = V_1(f_r) - V_1(e_r)$, where $V_1(z_0)$, $z_0 = e_r$ or f_r, equals

$$
\begin{aligned}
&C(x,d_q,z_0) - C(a_p,d_q,z_0) - C(x,y,z_0) + C(a_p,y,z_0) \\
&= \int_0^1 \left(A_t\left(F_x^{\alpha}(t), \eta_q, c\right) - A_t\left(\mu_p^*, \eta_q, c\right) - A_t\left(F_x^{\alpha}(t), F_y^{\beta}(t), c\right) + A_t\left(\mu_p^*, F_y^{\beta}(t), c\right)\right) \mathrm{d}t \\
&= \left[\int_0^{\alpha(x)} A_t\left(\mu_p, \eta_q, c\right) \mathrm{d}t + \int_{\alpha(x)}^1 A_t\left(\mu_p^*, \eta_q, c\right) \mathrm{d}t \right] - \int_0^1 A_t\left(\mu_p^*, \eta_q, c\right) \mathrm{d}t \\
&\quad - \left[\int_0^{\alpha(x)} A_t\left(\mu_p, \eta_q, c\right) \mathrm{d}t + \int_{\alpha(x)}^{\beta(y)} A_t\left(\mu_p^*, \eta_q, c\right) \mathrm{d}t + \int_{\beta(y)}^1 A_t\left(\mu_p^*, \eta_q^*, c\right) \mathrm{d}t \right] \\
&\quad + \left[\int_0^{\beta(y)} A_t\left(\mu_p^*, \eta_q, c\right) \mathrm{d}t + \int_{\beta(y)}^1 A_t\left(\mu_p^*, \eta_q^*, c\right) \mathrm{d}t \right] = 0,
\end{aligned}
$$

where $c = F_{z_0}^{\gamma}(t) = \begin{cases} \sigma_r^* & \text{if } z_0 = e_r, \\ \sigma_r & \text{if } z_0 = f_r. \end{cases}$ Hence, $V_C(\pi_{XY}(B) \times K_r) = 0$ i.e. $V_C(B) = 0$.

To prove Theorem 2 (2), some lemmas are in order. Their proofs are similar to the two dimensional version in [9] and hence omitted.

Lemma 1. *Let $\alpha, \beta, \gamma \in \mathscr{T}_P$ with the partitions $\{I_n\}, \{J_n\}$ and $\{K_n\}$, respectively, and $C \in \mathscr{C}_{PID}$ corresponding to α, β, γ. Let $i, q, r \in \mathbb{N}$, and $y, z \in (0,1)$. Then*

$$
K_C\left(s_{(i)}, (0, y^{(q)}] \times (0, z^{(r)}]\right) = K_C\left(s_{(i)}, \bigcup_{j \in \mathcal{J}} \bigcup_{k \in \mathcal{K}} (0, \beta_j^{-1}\alpha_i(s_{(i)})] \times (0, \gamma_k^{-1}\alpha_i(s_{(i)})]\right)
$$

for a.e. s, where $\mathcal{J} := \begin{cases} Q_q; & \beta_q^{-1}\alpha_i(s_{(i)}) \le y^{(q)}; \\ Q_q^*; & \beta_q^{-1}\alpha_i(s_{(i)}) > y^{(q)}, \end{cases}$ $\mathcal{K} := \begin{cases} R_r; & \gamma_r^{-1}\alpha_i(s_{(i)}) \le z^{(r)}; \\ R_r^*; & \gamma_r^{-1}\alpha_i(s_{(i)}) > z^{(r)}, \end{cases}$ $Q_q := \{j : J_j \preceq J_q\}$, $Q_q^* := \{j : J_j \prec J_q\}$, $R_r := \{k : K_k \preceq K_r\}$ *and* $R_r^* := \{k : K_k \prec K_r\}$.

For almost every $s \in \mathbb{I}$, we let $t, u \in \mathbb{I}$ be such that $\alpha(s) = \beta(t) = \gamma(u)$. Then we define $P(s) := \{0,1\} \cup \{\mu_n, \mu_n^*\}_{n\in\mathbb{N}}$, $Q(s) := \{0,1\} \cup \{\eta_n, \eta_n^*\}_{n\in\mathbb{N}}$ and $R(s) := \{0,1\} \cup \{\sigma_n, \sigma_n^*\}_{n\in\mathbb{N}}$ where each symbol follows from the notations below Theorem 2. In addition, from the totally ordered property of $P(s)$, every $\mu \in P(s)$ can be written as $\mu = \sum_{i\in\mathcal{I}_\mu} \frac{1}{|\alpha'(s_{(i)})|}$ for some $\mathcal{I}_\mu \in \{\{i : I_i \prec I_k\}, \{i : I_i \preceq I_k\} : k \in \mathbb{N}\}$. The elements $\eta \in Q(s)$ and $\sigma \in R(s)$ can be written in similar ways.

Lemma 2. *Let* $C \in \mathscr{C}_{PID}$ *corresponding to* $\alpha, \beta, \gamma \in \mathscr{T}_P$. *For each* $s \in I_1$, *define*

$$A_{\alpha(s)}(\mu,\eta,\sigma) := \sum_{i\in\mathcal{I}_\mu} \frac{1}{|\alpha_i'(s_{(i)})|} K_C\left(s_{(i)}, \bigcup_{j\in\mathcal{J}_\eta}\bigcup_{k\in\mathcal{K}_\sigma} [0,\beta_j^{-1}\alpha_1(s)] \times [0,\gamma_k^{-1}\alpha_1(s)]\right). \tag{3}$$

Then $A_{\alpha(s)}$ *is a 3-subcopula on* $P(s) \times Q(s) \times R(s)$ *extensible to a 3-copula.*

Proof (of Theorem 2 (2)). Putting $\mathcal{A} := \{A_t\}_{t\in\mathbb{I}}$ as defined in (3), it can be shown that $A_t(\partial_2 C_{e,\alpha}(x,t), \partial_2 C_{e,\beta}(y,t), \partial_2 C_{e,\gamma}(z,t))$ is measurable in $t \in \mathbb{I}$ and

$$\left(\underset{\mathcal{A}}{*}(C_{e,\alpha}, C_{e,\beta}, C_{e,\gamma})\right)(x,y,z) = \int_{a_1}^{b_1} |\alpha'(s)| A_{\alpha(s)}(u(s), v(s), w(s))\, \mathrm{d}s, \tag{4}$$

where $u(s) := \partial_2 C_{e,\alpha}(x, \alpha(s))$, $v(s) := \partial_2 C_{e,\beta}(y, \alpha(s))$, $w(s) := \partial_2 C_{e,\gamma}(z, \alpha(s))$.

Denote $C := C_{(X,Y,Z)}$ and K_C its Markov kernel. By Remark 1 (3), if $(x,y,z) \in I_p \times J_q \times K_r$ then, with $\{\hat{\mu}_p, \check{\mu}_p\} = \{\mu_p, \mu_p^*\}$, $\{\hat{\eta}_q, \check{\eta}_q\} = \{\eta_q, \eta_q^*\}$ and $\{\hat{\sigma}_r, \check{\sigma}_r\} = \{\sigma_r, \sigma_r^*\}$, for a.e. $s \in I_1$,

$$\begin{aligned} u &= \hat{\mu}_p \mathbb{1}_{(a_1, x_{(1)})} + \check{\mu}_p \mathbb{1}_{(x_{(1)}, b_1)}, \quad v = \hat{\eta}_q \mathbb{1}_{(a_1, \alpha_1^{-1}\beta_q(y))} + \check{\eta}_q \mathbb{1}_{(\alpha_1^{-1}\beta_q(y), b_1)} \\ w &= \hat{\sigma}_r \mathbb{1}_{(a_1, \alpha_1^{-1}\gamma_r(z))} + \check{\sigma}_r \mathbb{1}_{(\alpha_1^{-1}\gamma_r(z), b_1)} \end{aligned} \tag{5}$$

where $\hat{\mu}_p := \begin{cases} \mu_p & \text{if } \alpha_p^{-1}\alpha_1 \text{ is increasing;} \\ \mu_p^* & \text{if } \alpha_p^{-1}\alpha_1 \text{ is decreasing} \end{cases}$ and $\hat{\eta}_q, \hat{\sigma}_r$ are similarly defined according to the monotonicity type of $\beta_q^{-1}\alpha_1, \gamma_r^{-1}\alpha_1$, respectively.

The integral on I_1 in (4) will be split into four subintervals depending on the locations of $x_{(1)}, \alpha_1^{-1}\beta_q(y)$ and $\alpha_1^{-1}\gamma_r(z)$. For lack of space, we shall consider only the case when $x_{(1)} \le \alpha_1^{-1}\beta_q(y) \le \alpha_1^{-1}\gamma_r(z)$. By (5), on each of the subintervals $T_1 := [a_1, x_{(1)}]$, $T_2 := [x_{(1)}, \alpha_1^{-1}\beta_q(y)]$, $T_3 := [\alpha_1^{-1}\beta_q(y), \alpha_1^{-1}\gamma_r(z)]$ and $T_4 := [\alpha_1^{-1}\gamma_r(z), b_1]$, the integrand equals $|\alpha'(s)| A_{\alpha(s)}(\vec{\nu}_l)$ where $\vec{\nu}_1 := (\hat{\mu}_p, \hat{\eta}_q, \hat{\sigma}_r)$, $\vec{\nu}_2 := (\check{\mu}_p, \hat{\eta}_q, \hat{\sigma}_r)$, $\vec{\nu}_3 := (\check{\mu}_p, \check{\eta}_q, \hat{\sigma}_r)$ and $\vec{\nu}_4 := (\check{\mu}_p, \check{\eta}_q, \check{\sigma}_r)$. That is,

$*_{\mathcal{A}}(C_{e,\alpha}, C_{e,\beta}, C_{e,\gamma}) = \sum_{l=1}^{4} \int_{T_l} |\alpha'(s)| A_{\alpha(s)}(\vec{\nu}_l)\mathrm{d}s$. By (3), each integral equals $\int_{T_l} \sum_i |\alpha'_{i1}(s)| K_C\left(s_{(i)}, \bigcup_j \bigcup_k B_{\beta_j^{-1}\alpha(s), \gamma_k^{-1}\alpha(s)}\right) \mathrm{d}s$ where the indices i, j, k range over sets corresponding to $\vec{\nu}_l$ and $B_{a,b} := [0,a] \times [0,b]$. For instance, for $s \in T_3$ in which $\vec{\nu}_3 = (\check{\mu}_p, \check{\eta}_q, \hat{\sigma}_r)$, i, j, k range over $\mathcal{I}_{\check{\mu}_p}, \mathcal{J}_{\check{\eta}_q}, \mathcal{K}_{\hat{\sigma}_r}$, respectively.

By Remark 1(2), if $\beta_q^{-1}\alpha_1$ and $\gamma_r^{-1}\alpha_1$ are increasing, then for $s \in T_1 \cup T_2$, $s_{(i)} = \alpha_{i1}(s) < \alpha_{i1}\alpha_1^{-1}\beta_q(y) = \alpha_i^{-1}\beta_q(y)$ and $s_{(i)} < \alpha_i^{-1}\gamma_r(z)$ if α_{i1} is increasing or $s_{(i)} > \alpha_i^{-1}\beta_q(y)$ and $s_{(i)} > \alpha_i^{-1}\gamma_r(z)$ if α_{i1} is decreasing. Both situations lead to $\beta_q^{-1}\alpha_i(s_{(i)}) < y$ and $\gamma_r^{-1}\alpha_i(s_{(i)}) < z$. Thus by Lemma 1, $K_C\left(s_{(i)}, \bigcup_{j\in\mathcal{J}_{\check{\eta}_q}} \bigcup_{k\in\mathcal{K}_{\hat{\sigma}_r}} B_{\beta_j^{-1}\alpha(s),\gamma_k^{-1}\alpha(s)}\right) = K_C\left(s_{(i)}, B_{y,z}\right)$. Similar results hold for other values of s. By the changes of variable $s_{(i)} = \alpha_{i1}(s)$ for $i \in \mathcal{I}_{\mu_p}$, $*_{\mathcal{A}}(C_{e,\alpha}, C_{e,\beta}, C_{e,\gamma})(x,y,z)$ equals

$$= \left(\sum_{\substack{i\in\mathcal{I}_{\hat{\mu}_p} \\ \alpha_{i1}\text{ inc}}} \int_{a_i}^{x_{(i)}} + \sum_{\substack{i\in\mathcal{I}_{\hat{\mu}_p} \\ \alpha_{i1}\text{ inc}}} \int_{x_{(i)}}^{b_i} + \sum_{\substack{i\in\mathcal{I}_{\hat{\mu}_p} \\ \alpha_{i1}\text{ dec}}} \int_{x_{(i)}}^{b_i} + \sum_{\substack{i\in\mathcal{I}_{\hat{\mu}_p} \\ \alpha_{i1}\text{ dec}}} \int_{a_i}^{x_{(i)}} \right) K_C(s, B_{y,z})\mathrm{d}s$$

$$= \sum_{\substack{i\in\mathcal{I}_{\mu_p^*} \\ \alpha_{i1}\text{ inc}}} \int_{a_i}^{b_i} K_C\left(s, B_{y,z}\right)\mathrm{d}s + \sum_{\substack{i\in\mathcal{I}_{\mu_p^*} \\ \alpha_{i1}\text{ dec}}} \int_{a_i}^{b_i} K_C\left(s, B_{y,z}\right)\mathrm{d}s + \int_{a_p}^{x} K_C\left(s, B_{y,z}\right)\mathrm{d}s$$

which, by the assumption on α_{p1}, equals to $\int_0^x K_C(s, B_{y,z})\mathrm{d}s = C(x,y,z)$.

Acknowledgements. We would like to thank the reviewer for comments and suggestions which considerably improve the readability of the manuscript.

References

1. Ansari, J., Rüschendorf, L.: Sklar's theorem, copula products, and ordering results in factor models. Depend. Model. **9**, 267–306 (2021)
2. Balakrishnan, N., Lai, C.D.: Continuous Bivariate Distributions, 2nd edn. Springer Science+Business Media (2009)
3. Darsow, W.F., Olsen, E.T.: Characterization of idempotent 2-copulas. Note Mat. **30**(1), 147–177 (2010)
4. de Amo, E., Díaz Carrillo, M., Fernández-Sánchez, J.: Measure-preserving functions and the independence copula. Mediterr. J. Math. **8**, 431–450 (2011)
5. Durante, F., Klement, E.P., Quesada-Molina, J.J., Sarkoci, P.: Remarks on two product-like constructions for copulas. Kybernetika **43**(2), 235–244 (2007)
6. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press (2015)
7. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer Science+Business Media, New York (2006)
8. Panyasakulwong, P., Santiwipanont, T., Sumetkijakan, S.: Generalized factorizability of some implicit dependence copulas. In: Proceedings of the Annual Pure and Applied Mathematics Conference, pp. 68–85. Chulalongkorn University, Thailand (2021)
9. Panyasakulwong, P., Santiwipanont, T., Sumetkijakan, S., Yanpaisan, N.: Countably piecewise monotonic surjection implicit dependence copulas. Fuzzy Sets Syst. **448**, 49–64 (2022)

10. Rényi, A.: Probability Theory. North-Holland, Amsterdam (1970)
11. Ruankong, P., Sumetkijakan, S.: On a generalized *-product for copulas. In: Proceedings of the Annual Pure and Applied Mathematics Conference, pp. 13–21. Chulalongkorn University, Thailand (2011)
12. Schreyer, M., Paulin, R., Trutschnig, W.: On the exact region determined by Kendall's τ and Spearman's ρ. J. R. Stat. Soc. B **79**(2), 613–633 (2017)
13. Sumetkijakan, S.: Non-atomic bivariate copulas and implicitly dependent random variables. J. Statist. Plann. Inference **186**, 1–12 (2017)
14. Trutschnig, W.: Complete dependence everywhere? In: Copulas and Dependence Models with Applications: Contributions in Honor of R.B. Nelsen, pp. 225–240. Springer International Publishing (2017)
15. Trutschnig, W.: On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. **384**(2), 690–705 (2011)
16. Vitale, R.A.: Parametrizing doubly stochastic measures. IMS Lecture Notes Monogr. Ser. **28**, 358–364 (1996)

Python-Based Reinforcement Learning on Simulink Models

Georg Schäfer[1,2,3(✉)], Max Schirl[2,3], Jakob Rehrl[1,2], Stefan Huber[1,2], and Simon Hirlaender[3]

[1] Josef Ressel Centre for Intelligent and Secure Industrial Automation, Salzburg, Austria
[2] Salzburg University of Applied Sciences, Salzburg, Austria
[3] Paris Lodron University of Salzburg, Salzburg, Austria
georg.schaefer@fh-salzburg.ac.at

Abstract. This paper proposes a framework for training Reinforcement Learning agents using Python in conjunction with Simulink models. Leveraging Python's superior customization options and popular libraries like Stable Baselines 3, we aim to bridge the gap between the established Simulink environment and the flexibility of Python for training bleeding edge agents. Our approach is demonstrated on the Quanser Aero 2, a versatile dual-rotor helicopter. We show that policies trained on Simulink models can be seamlessly transferred to the real system, enabling efficient development and deployment of Reinforcement Learning agents for control tasks. Through systematic integration steps, including C-code generation from Simulink, DLL compilation, and Python interface development, we establish a robust framework for training agents on Simulink models. Experimental results demonstrate the effectiveness of our approach, surpassing previous efforts and highlighting the potential of combining Simulink with Python for Reinforcement Learning research and applications.

Keywords: Reinforcement learning · Simulink · Python · Stable Baselines 3

1 Introduction

Reinforcement Learning (RL) has recently gained much interest in the field of control of dynamical systems [2,5], as RL is a valuable alternative to traditional and advanced control concepts. In many mechatronic control tasks, performance can be considerably improved by using these advanced concepts over standard controllers like PID. Some examples are linear quadratic regulators or linear Model Predictive Control (MPC). However, for nonlinear systems, these concepts may not exploit the full potential of the mechatronic system. Although

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 449–456, 2024.
https://doi.org/10.1007/978-3-031-65993-5_55

nonlinear control approaches exist, their development is time-consuming, since a concept tailored to the specific nonlinear system needs to be created. The computational effort of nonlinear MPC renders real-time execution of the control law impossible for many applications. RL can overcome some of these limitations. The nonlinear system behavior can be taken into account at no additional cost, and the execution of the (greedy) agent is often computationally lightweight.

Often, the design of the RL policy is not done directly using the real environment, but instead using a simulation model of the real system that mimics the environment. A common toolset for setting up the simulation model is MATLAB/Simulink. Specifically, the realization of the simulation model by means of block diagrams with the help of Simulink is an effective and well established workflow. However, if developing bleeding edge RL agents is the goal, Python may be seen as the de facto standard in the industry. OpenAI Baselines, Stable Baselines3 (SB3), RLlib, TF-Agents among other sets of implementations are only available in Python and offer superior customization options when it comes to the creation of custom agents. In the present paper, a framework combining these two programming languages is proposed. For that purpose, a lab-scale model of a mechatronic system is utilized: the Quanser Aero 2.

1.1 Motivation

Our goal is to train a RL policy using Python, more specifically relying on SB3 [8], and a simulated environment of the Quanser Aero 2 based on MATLAB/Simulink. We also want to show that this policy can be transferred to the real system. In the end, the generated policy is therefore not only able to control the simulated environment based on Simulink, but also the real system.

This specific test bed, the aforementioned Aero 2, is a versatile dual-rotor helicopter and is shown in Fig. 1. The system is interesting for RL related research since it

- supports multiple inputs and outputs (MIMO),
- is a non-linear system,
- may be configured in different ways concerning its axes' Degrees of freedom (DOF) and
- provides a system model, allowing training in a (Simulink) simulation.

Using a test bed like this represents industry standards, as simulations and mechatronical models of real systems are available in MATLAB/Simulink. This work aims to expand upon this standard by introducing Python-based RL. More specifically, starting out in a 1-DOF configuration of the Quanser Aero 2, allows us to train a RL policy on a simple balancing task, evaluating the feasibility of Python frameworks for RL tasks. In future research, learning more complex balancing tasks (with 2-DOF) can be achieved by simply changing the settings of the real system and its simulation.

1.2 Related Work

Fandel et al. [1] use the Quanser Aero in a 2-DOF configuration to solve a balancing task using RL based on MATLAB. This paper shows that, in prin-

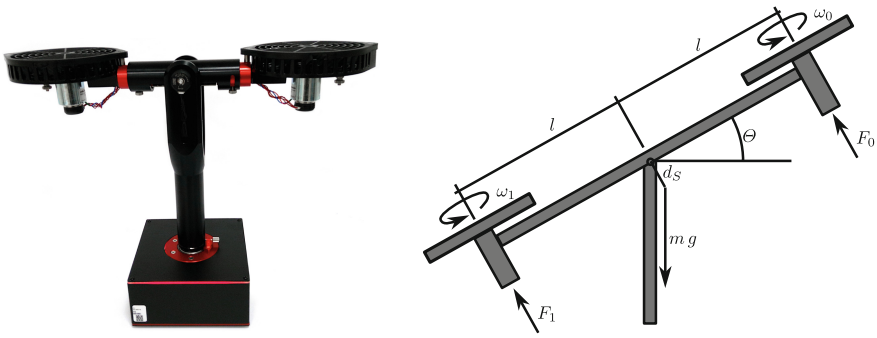


Fig. 1. The Quanser Aero 2 (left) and its schematic representation (right) in a 1-DOF configuration.

ciple, training RL agents using the MATLAB ecosystem in connection with Quanser's devices is feasible [1]. Polzounov et al. [6] present a similar approach that substitutes Simulink with a self-developed Python-based driver to simulate the environment. The work presented is based on a Qube–Servo 2 system by Quanser and focuses on training RL agents on the real system [6]. In [9], the authors compare the performance of fine-tuned RL agents to the performance of MPC-based controllers in a 1-DOF balancing task using the Aero 2. The work is purely based on MATLAB and shows that the default configuration of the Proximal Policy Optimization (PPO) agent in MATLAB is insufficient to solve this basic task without fine-tuning. Even tough the complexity of the task was reduced to a minimum, custom parametrization was needed. This is unexpected, as PPO is known for its stability.

1.3 Main Contribution

In essence, we aim to expand upon the approach of [6], moving away from MATLAB and onto Python when training RL agents, while relying on interfaces provided by Gymnasium [10] to connect to Simulink environments. This allows us to integrate the models created via Simulink into the RL ecosystems provided by popular Python frameworks such as SB3, without the need to hand-craft equations that determine the system's behavior.

1.4 Reinforcement Learning Setting

The task to be solved consists of setting the Aero 2's beam angle (Θ) to a specific angle (r) by adjusting the voltage applied to its motors (u and $-u$). The action u is updated every 0.1 s to achieve this goal. The system's state representation consists of the distance to the desired angle ($\Delta = \Theta - r$) and the current angular velocity ($\omega = \dot{\Theta}$). The reward function is defined as the negative absolute distance to the target angle ($-|\Delta|$), motivating the agent to minimize the deviation from the desired orientation. This is the same setting that is used

in [9] and therefore allows the comparison of the performance of agents shown in the present paper to the results presented [9].

2 Approach

We present a systematic approach for integrating Simulink models with Python RL implementations. The process involves four main steps:

1. **Generating C-Code from Simulink**: By utilizing Simulink Coder [4], the control model can be translated to C-Code.
2. **Compiling to Dynamic Link Library (DLL)**: The generated C-Code is compiled into a DLL, thus enabling seamless integration.
3. **Embedding DLL in Python**: The DLL is interfaced with Python using ctypes, exposing the following essential functions and data structures:
   - `MODELNAME_initialize`: Initializes the control model.
   - `MODELNAME_step`: Performs a single step of the control model.
   - `MODELNAME_terminate`: Terminates the control model.
   - `MODELNAME_U`: Accesses input data.
   - `MODELNAME_Y`: Retrieves output data.
4. **Creating Gymnasium Environment:** A custom Gymnasium [10] environment is constructed to interface with the Simulink model, implementing the following key methods (see Code 55.1 in the Appendix for details):
   - `__init__`: Defines the action and observation spaces, and initializes the control model.
   - `step`: Executes an action, updates the control model, and returns the new state and reward. To comply with the agent's specified sample time, the model update must be repeated based on the Simulink model's configured step-size.
   - `reset`: Resets the environment, reinitializes the control model, and clears input data.
   - `render`: Provides optional visualization for monitoring system behavior.

Using the created custom Gymnasium environment, a policy can be identified by employing state-of-the-art Python RL agent implementations.

2.1 Deployment

To be able to deploy the trained policy to the real system, the Quanser Python API [7] can be utilized alongside the provided Hardware-in-the-Loop (HIL) card. As the pitch Θ is needed to define both states ($\Delta = \Theta - r$) and ($\omega = \dot{\Theta}$), the pitch encoder needs to be used to read the aforementioned pitch. ω is determined using a second order low-pass filter, as the first derivative of Θ can not be captured directly from the real system. By relying on these two states, the greedy action u can be determined using the identified policy learned on the simulated system. In accordance with this action, the voltages u and $-u$ are applied to the corresponding motors by directly writing to the analog channels of the HIL card. To comply with the sample time of the agent, a software timer is employed, triggering actions accordingly.

3 Discussion

To solve the problem formulated in Sect. 1.4, the SB3's PPO agent implementation with default parameterization was used. A dynamic tilt function, as illustrated in Fig. 3, was used to determine the target pitch r. Each training run took 80 s, with changes in target pitch occurring every 10 s. Figure 2 illustrates the mean episode return of five independent runs for SB3's PPO and MATLAB's PPO implementation in the simulation where each run's first 720.000 steps are displayed. With the agent's sample time set to 0.1 s, each episode takes 800 steps, resulting in 900 episodes per training run. The highest return reached for SB3 is -64.87, representing a mean average deviation to the target tilt of 4.6°. In contrast, MATLAB's PPO agent reached a maximum return of -77.93 (average distance to target of 5.6°) after fine-tuning parameters as shown in [9]. To reiterate, SB3 achieved a better result without the need to fine-tune hyperparameters.

Figure 3 shows the evaluation of the agent trained on the simulated model in a real world scenario alongside its performance in the simulation. It is evident that PPO is able to solve this task adequately in the simulation environment, and shows promise when applied to the real world system, when keeping in mind that the agent was solely trained on the simulation and therefore had no access to the real world system during training time. Providing an optimal solution to the balancing task, that is solving the problem of steady state deviations as well as oscillations present for both Θ_{sim} and Θ_{real}, is decidedly not in the scope of this work, but is part of our future research.

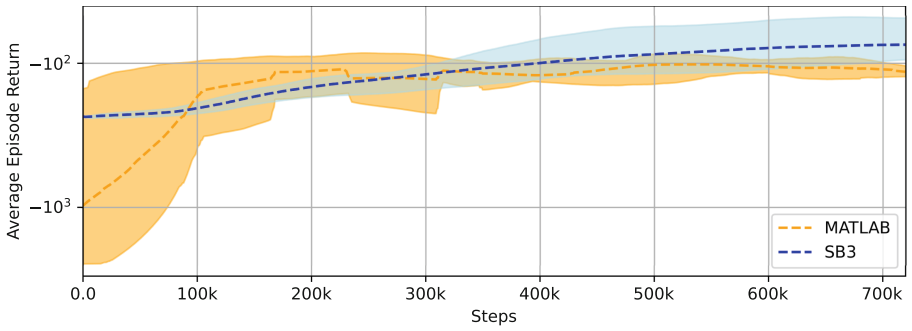


Fig. 2. A comparison of the average episode return of SB3's PPO agent (in blue) versus MATLAB's PPO implementation (in orange) on a logarithmic y-axis. The dashed lines represent mean episode return of five training runs each, with the areas in blue and orange representing minimum and maximum episode return, respectively.

4 Conclusion and Future Work

To summarize, we show that training RL policies on the basis of a simulated representation of Quanser's Aero 2 using Simulink in combination with Python

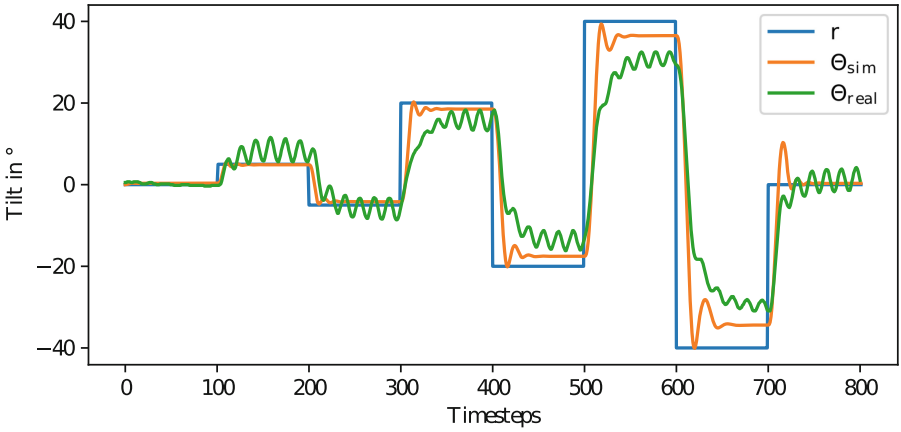


Fig. 3. Evaluation runs on the simulation and real system illustrating the behavior of the actual tilt Θ_{sim} and Θ_{real} for simulation and real system, respectively, using a greedy policy and a dynamic target tilt r.

frameworks such as SB3 is feasible. Additionally, we have shown that the learned policy is transferable to the real system. Most importantly, the default configuration of SB3's PPO implementation outperforms the fine-tuned PPO agent based on MATLAB's RL Toolbox presented in [9]. Note that the default configuration of MATLAB's PPO agent was not able to solve the balancing task at all. Being able to rely on the default settings for the PPO agent without having to fine-tune hyperparameters significantly decreased development time.

In future work, we want to compare the implementations of SB3 and MATLAB's RL Toolbox [3] in more detail. Additionally, fine-tuning the SB3 PPO agent may lead to increased performance exhibited in the balancing task for both the simulated environment and the real world system. Two additional ways to improve performance for the real system include the expansion of the state space, by adding the actual tilt Θ, and the training of the PPO agent on the real system itself. Other planned areas of research include the comparison of the quality of the learned policies balancing capabilities to default controllers such as linear-quadratic regulators or model predictive control. Furthermore, the 2-DOF configuration of the Quanser Aero 2 system will be considered in future work.

Acknowledgments. Financial support for this study was provided by the Christian Doppler Association (JRC ISIA), the WISS-FH project IAI, the European Interreg Österreich-Bayern project BA0100172 AI4GREEN and the PRISMATICS research project.

Appendix

The following listing shows the primary implementation for generating a Gymnasium environment utilizing a DLL generated by Simulink Coder in the specific example of the Quanser Aero 2 system.

```
import gymnasium as gym
import ctypes
import numpy as np

class AeroEnv(gym.Env):
def __init__(self):
self.action_space = gym.spaces.Box(
low=-24.0, high=24.0, shape=(1,)
)
self.observation_space = gym.spaces.Box(
low=-np.pi, high=np.pi, shape=(2,)
)
self.model = ctypes.CDLL("./aero.dll")
self.input = get_input(self.model, "aero_U")
self.output = get_output(self.model, "aero_Y")
self.model.aero_initialize()

def step(self, action):
self.input.v0 = action[0]
self.input.v1 = action[1]
for _ in range(5):
self.model.aero_step()

return self.state, self.reward, False, False, {}

def reset(self):
self.model.aero_terminate()
self.model.aero_initialize()
self.input.v0 = 0.0
self.input.v1 = 0.0
return self.state, {}

@property
def state(self):
return np.array(
[self.target_tilt - self.output.pitch, self.output.
velocity]
)

@property
def reward(self):
return -np.abs(self.target_tilt - self.output.pitch)
```

Code 55.1. Custom Gymnasium Environment for the Aero 2

References

1. Fandel, A., Birge, A., Miah, S.: Development of reinforcement learning algorithm for 2-dof helicopter model. In: 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE), pp. 553–558. IEEE (2018)
2. Li, Y.: Deep Reinforcement Learning: An Overview. arXiv preprint arXiv:1701.07274 (2017)
3. MathWorks: Agents—Matlab & Simulink. https://de.mathworks.com/help/reinforcement-learning/agents.html (visited on 16 Mar 2024)
4. MathWorks: Simulink Coder—Matlab. https://de.mathworks.com/products/simulink-coder.html (visited on 16 Mar 2024)
5. Padakandla, S.: A survey of reinforcement learning algorithms for dynamically varying environments. ACM Comput. Surv. **54**(6), 1–25 (2021)
6. Polzounov, K., Sundar, R., Redden, L.: Blue River Controls: A Toolkit for Reinforcement Learning Control Systems on Hardware. arXiv preprint arXiv:2001.02254 (2020)
7. Quanser: Quanser Python API Documentation. https://docs.quanser.com/quarc/documentation/python/index.html (visited on 16 Mar 2024)
8. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N.: Stable-baselines3: Reliable reinforcement learning implementations. J. Mach. Learn. Res. 22(268), 1–8 (2021). http://jmlr.org/papers/v22/20-1364.html
9. Schäfer, G., Rehrl, J., Huber, S., Hirlaender, S.: Exploring the Dynamics of Reinforcement Learning in Aerospace Control. Presented at the RL4AA Salzburg (2024)
10. Towers, M., Terry, J.K., Kwiatkowski, A., Balis, J.U., Cola, G., Deleu, T., Goulão, M., et al.: Gymnasium (v0.28.1). Zenodo (2023). https://doi.org/10.5281/zenodo.8127026 (visited on 8 July 2023)

Applied Statistics in Industry: Defining an Appropriate Target Variable and Analysing Factors Affecting Aluminium Ingot Quality

Manuela Schreyer[1(✉)], Marco Tschimpke[2], Alexander Gerber[3], Steffen Neubert[3], and Wolfgang Trutschnig[2]

[1] Austria Metall GmbH, Lamprechtshausenerstr. 61, 5282 Ranshofen, Austria
manuela.schreyer@amag.at
[2] Department for Artificial Intelligence and Human Interfaces, University of Salzburg, Hellbrunner Straße 34, 5020 Salzburg, Austria
[3] AMAG Casting GmbH, Lamprechtshausenerstr. 61, 5282 Ranshofen, Austria

Abstract. Methods from the field of data science are becoming increasingly important in various areas of the economy. One of the main applications in industry is the continuous improvement of product quality, which is becoming increasingly challenging due to growing quality demands. A variety of statistical and machine learning methods are available for this purpose. However, to use these methods for a specific application, it is necessary to define an appropriate target variable. This can be a significant challenge due to the extensive product portfolio. The target variable must cover many cases and remain comparable for all cases under consideration. Limiting it to a single case would often result in a database too small for any analysis. This work aims at improving the quality of aluminium plates for the aerospace industry by analysing casting process data. The focus is on the development of a target value for this issue. Additionally, a brief outlook is given on the subsequent analysis of possible influencing variables and the development of a predictive model of product quality.

Keywords: Target variable · Distribution · Prediction · Machine learning · Statistical methods · Aluminium direct chill casting

1 Introduction

The use of data science is becoming more and more important in many areas due to increasing quality requirements. However, the sheer volume and complexity of production data requires effective methods to extract insights from data to further optimise production processes. This work considers production data from the aluminium industry, in particular from aluminium direct chill casting. Up to six ingots are cast in one batch from the input material, which are then

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 457–465, 2024.
https://doi.org/10.1007/978-3-031-65993-5_56

rolled, and finally cut into plates. Throughout the melting and casting process, hundreds of sensor measurements are continuously taken for several hours. After casting, the entire rolling process is carried out over a period of several weeks. Finally, the material is inspected at the end of the production process. The final product must fulfil high quality requirements depending on the customer's specifications. Strict quality controls ensure that these requirements are met [7].
To further improve product quality, it is essential to analyse process data and draw conclusions about product quality. This enables early measures to be taken to avoid rejections and reduce scrap, delivery times, plant utilisation and costs in general. To achieve this goal, various methods from statistics and machine learning can be used. However, in order to apply any of these methods, it is necessary to define an appropriate target variable.
AMAG is the world's only producer of all aluminium alloys (1xxx–8xxx) including most casting alloys at one location. In order to have a database as large as possible for analysis, it is necessary to define a target variable that is comparable for the widest possible selection of different alloys, casting formats, plate formats, systems, etc. This is one of the major challenges of data analysis in this environment. If the database were limited to a single case, the database would be too small for any analysis.
This article describes the methodology used to calculate the target variable for a specific application in the aluminium industry using machine learning algorithms and statistical methods.

2 Application Example

During the casting process of rolling ingots, non-metallic inclusions may occur. Identifying potential factors that influence the casting process and cause quality deviations allows for early actions (during/before casting) [6,10,11]. In this way, scrap, plant utilisation and delivery times can be reduced.
In the first step of the production process, the input material is melted, which could be the first possible source of quality deviations. The subsequent melting and casting process is another possible source. After casting, the entire rolling process follows, which includes rolling, heat treatment, stretching, cutting, etc. At the end, a final quality check is carried out on the plates. Among other things, the sheets are ultrasonically (US) tested.
During US-inspection, the exact locations and sizes of non-metallic inclusions are determined in each plate. To identify potential factors that influence the formation of non-metallic inclusions during the casting process, it is necessary to develop a target variable describing the quality of the ingot directly after casting, independent of any further processing. Therefore, in a first step the positions of the indications in the plates are reverse engineered back to the ingot. The main challenge here is to consider the various sawing waste and large dimensional changes between process steps at different temperatures and therefore different densities. Due to the large dimensional changes, even a small deviation can result in a significant error when recalculating the inclusion positions.

Subsequently, the calculation of the target variable can be tackled. There are many challenges for a data scientist to overcome. Details on several of these aspects are given below. These are followed by a discussion of solutions to these issues. After this, the paper gives a brief overview of the associated development of a prediction model.
The target variable is determined initially for the primary rolling ingot format (thickness: 450 mm; with: 1400 mm; length: approximately 7000 mm) and for the three main alloys of the 7xxx alloy family, namely A1, A2, and B. The only difference between A1 and A2 is that A2 meets narrower tolerance windows for chemical composition.

2.1 Some Issues to Be Considered When Defining a Target Variable

2.1.1 Untested Areas in Length Direction In Fig. 1 the recalculated indications are depicted. The US-tested areas in length direction are shaded in dark grey. The upper and lower sections of the ingot are removed right after casting (shaded in red). They can vary significantly. Additional scrap of varying lengths is produced between the plates during the rolling process.

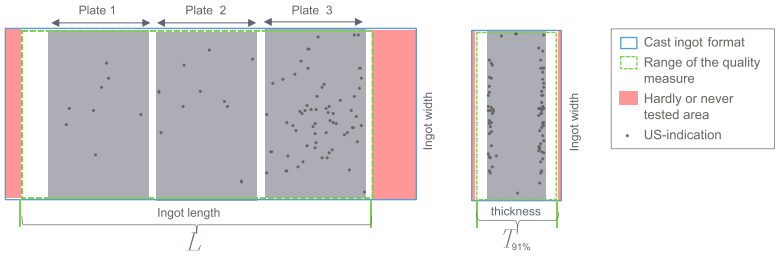


Fig. 1. Schematic representation

2.1.2 Untested Areas in Thickness Direction In addition to the untested areas in the longitudinal direction, there are also some in the thickness direction. Due to technical limitations, the outer 3mm at the top and bottom of the plate (in thickness direction) cannot be US tested. This area is milled away after the US-inspection by the customer as its quality cannot be determined (in the sequel referred to as 'US-dead zone'). The size of this area varies greatly depending on the plate. For instance, in the case of a plate with thickness of 25 mm, that area represents $(6/25 * 450)/450 = 24\%$ of an ingot with thickness of 450 mm. If the plate is 150 mm thick, it would only correspond to 4 % of the ingot.

2.1.3 Effect of Plate Thickness on Indication Size The thickness of the final plate affects not only the size of the area tested in the thickness direction,

but also the size of each indication detected by the US test. Depending on the final plate thickness, the rolling pressures on the ingot can be very different. This results in differences in the size of indications measured in the final plates.

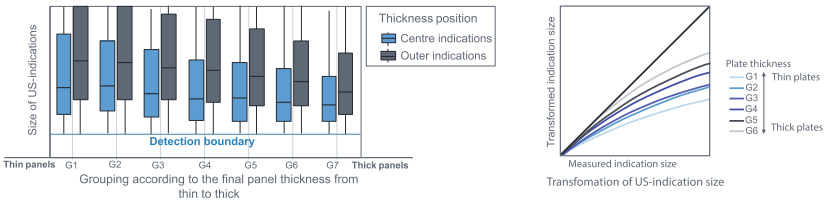


Fig. 2. Effects on indication size; alloy B

Figure 2 (left-hand side) depicts box plots of the indication sizes for alloy B, categorized by final plate thickness [8]. Group G1 on the x-axis includes the thinnest plates, while group G7 includes the thickest plates of this alloy. The indications were classified based on their thickness-position in the ingot as either 'Center indications', located in the center of the ingot, or 'Outer indications', located on the top and bottom of the ingot in thickness-direction. For both groups, it is observed that the size of the indication decreases with increasing plate thickness. The right-hand side of Fig. 2 shows the transformation of the US indications which is explained in below.

2.1.4 Effect of Indication Position in the Thickness Direction on Indication Size Different rolling pressures also occur in the thickness direction, especially for alloy B. As shown in Fig. 2 (left-hand side), 'outer indications' exhibit larger indication sizes compared to 'central indications'.

2.1.5 Different Indication Distribution in Thickness Direction Figure 3 depicts the distribution of the indications in thickness direction for the three alloys considered. The center of the ingot in thickness direction is referred to as the zero point on the x-axis. 225 mm represents the edges of the ingot. It should be noted that the outermost edges of an ingot are not US-tested for any plate format, as previously explained. However, all ingots are tested in the area to the left of the blue line for all plate thicknesses produced. The indication distribution also depends on the length position in the ingot, possibly due to the melting temperatures, which is part of the current research. Figure 3 illustrates the distribution of indications in thickness direction for indications along the entire length of the ingot.

Other factors that may affect the distribution of US-indications include casting temperature, chemical composition, and other rolling-related factors. A challenge in this context is the lack of prior knowledge on all influencing variables. Further research aims to identify additional factors that may affect the definition

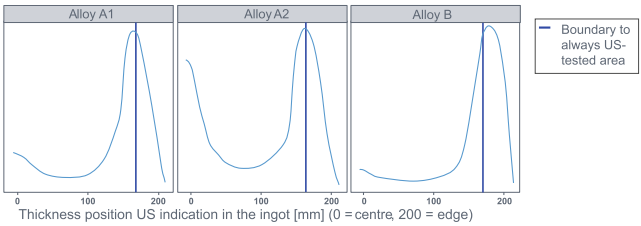


Fig. 3. Distribution of the indication position in thickness direction

of a target variable, and thus improve the analysis and the predictive models. The following section presents solutions for defining a target variable that is independent of the factors described above.

2.2 Approaches to Solutions

The descriptions for calculating a target variable below are based on the main format and alloy B.

During the US-test, a value is calculated describing the area (in mm^2) of the detected indication. Additionally, its position in the thickness, width, and length direction in the plate is recorded. The target variable should describe the sum of the indication areas [mm^2] per tested plate weight [t]. In the following it is also referred to as the quality measure defined by

$$QM_b := \frac{\sum_{i \in I_b} A_i}{W_b} \tag{1}$$

for an ingot b, where A_i, $i \in I_b = 1, \ldots, n_b$, is the indication area of the n_b indications in the ingot and W_b is the sum of the tested plate weights.

The target variable should describe the quality of the ingot in relation to US-indication, regardless of any subsequent process steps after casting. As previously mentioned, the US-tested area and indication size are significantly affected by rolling processes and the US-test methodology. To create a target variable that is independent of these processes, a series of transformations and estimations must be carried out.

Fig. 1 shows a schematic representation (no real data) of the different areas in the ingot, where three plates are produced from one ingot. The blue line illustrates the casting format of the ingot, including milling, sawing, and stretching waste. The dark grey areas indicate the regions that were scanned during the US-inspection of the three plates. Stretching and sawing areas are omitted as described previously. The size of this area can vary greatly and thus the area which is US-tested in length direction. QM_b always refers to the length interval in which 95 % of the ingots were tested (labelled with L in Fig. 1).

The same principle applies to the thickness direction. As previously stated, the immediate edges of each ingot are not tested due to the 'US-dead zone'. QM_b refers to the thickness interval in which a certain proportion of plates have been

produced (labelled with $T_{91\%}$ in Fig. 1). The interval $T_{91\%}$ describes 91 % of the ingot in thickness direction. This is the largest interval in thickness direction where plates have been tested, allowing for a reliable estimate of the quality measure, as will be shown later.
In width direction, always the same area of the ingot is US-tested. The green dashed line now illustrates the area $L \times T_{91\%}$ where QM_b of all ingots b should be calculated or estimated.
Marked in red is the area where the quality measure cannot be estimated because no (or hardly any) ingots are US-tested.

2.2.1 Estimation of US-Untested Areas The initial step is to estimate QM_b in the untested areas and to generate a measure QM_b that describes the ingot quality in the green marked areas in Fig. 1.

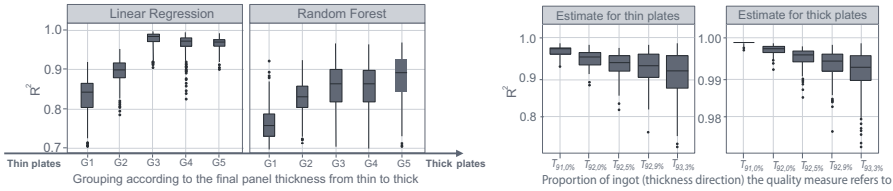


Fig. 4. Comparison of the performance of different models

To consider the distribution in thickness and length direction, the US indications were divided into several groups regarding ingot thickness and length. Various models were compared for different subdivisions in thickness and length directions [4,5]. Figure 4 (left-hand side) depicts some results (linear models and random forest) for estimations of QM_b of the whole area $L \times T_{91\%}$ for one specific subdivisions in thickness and length directions which showed the best result overall for alloy B.
Furthermore, the performance was tested not only for the interval $T_{91\%}$, which refers to 91 % of the ingot in the thickness direction, but also for other ranges. The results are presented in Fig. 4 (right-hand side) for the linear model, which exhibited the best performance. The diagram depicts the estimation for thin and thick plates. The untested area is significantly larger for thin plates than for thick plates, resulting in a less accurate estimate (note that the y-axes are scaled differently). The x-axis indicates the proportion of the ingot in thickness direction to which the quality measure pertains. It is evident that the estimate for $T_{91\%}$ also yields excellent outcomes for thin plates. However, the loss of information from $T_{93.3\%}$ to $T_{91\%}$ is relatively minor.

2.2.2 Elimination of Rolling Effects As demonstrated previously, the size of the indications strongly depends on their position in thickness direction as

well as on the final thickness of the plates due to different rolling forces. An effect of the indication position in length or width direction on the indication size was not observable. To eliminate this effect, the indication sizes are transformed before the estimation described above.

Figure 2 (left-hand side) shows that the size of the indication depends on both, the final plate thickness and the position of the indication in the ingot in the thickness direction. This indicates that the previously calculated quality measure is still influenced by the rolling process and thus the indication sizes are transformed beforehand.
For this reason, all indication sizes shown in Fig. 2 (left-hand side) of groups $G1, ..., G6$ are transformed in such a way that the corresponding distributions are similar to those of group $G7$ (i.e. minimum L_1-distance of the corresponding distribution functions; see [9]). Figure 2 (right-hand side) shows the transformation results (using 'optim' [8]), exemplary for indications in the center of the ingot in thickness direction, colored according to plate thickness group. It can be seen that the transformation for group G6 (light grey) is closest to $f(x) = x$ (shown in black) compared to the other groups. This is also consistent with the results shown in Fig. 2 (left-hand side).

2.2.3 Calculation of a Quality Measure per Batch During the casting process, hundreds of parameters are continuously measured. However, most of these parameters affect the entire batch. Previously, the calculation of a quality measure per ingot-section was described. In the following, a brief explanation is given on how to calculate a quality measure per batch.
A batch contains up to six ingots. It is possible for a single ingot within a batch to be of poor quality, even if the remaining ingots are of very good quality. This phenomenon can also affect individual ingot sections (hereafter referred to as an outlier ingot (sections)). Obviously, this cannot be caused by a batch related variable such as melt temperature. Instead, it must be an ingot-specific influencing variable that is measured separately for each ingot position.
To identify such outlier ingot sections, a Monte Carlo simulation is used to calculate the probability that any ingot section within a batch section of a total of N indications will have x indications more than the other ingot sections, assuming that the N indications of the entire batch section are evenly distributed across all ingot sections. If this probability is less than a specified value, the ingot section is classified as an outlier. To consider the different final plate thicknesses, the number of US indications was also estimated for each ingot in the untested area in a first step (procedure similar to the estimation of the quality measure).

3 Results and Further Work

The previous sections demonstrated the calculation of a target variable (quality measure) per ingot and per batch section. This variable can be used to develop a prediction model for ingot and batch quality. The model can assist technologists in gaining initial insights and making faster decisions about further processing. Prediction models for batch sections were trained using historical data of specific alloys over several years, and their performance was evaluated (results see [10,11]). While the current model has demonstrated promising results for a wide range, additional development is required before implementation in production. Additionally, potential variables that may influence product quality should be identified so that they can be countered at an early stage. Various methods were used to select features, resulting in the reduction of hundreds of available parameters to a few relevant ones [1,3,10,11]. Additionally, hypotheses developed by technologists based on technical aspects were analysed [2,11].

Both approaches enable the improvement of the production process. However, there are always unknown parameters and correlations that have not yet been identified. The prediction model aims to assist the technologist's decision making by providing a certain probability of rejection immediately after the ingot casting. It also significantly improves the understanding of the interactions between parameters and the overall production process.

In future work, the implementation of a prediction model will be pursued. A major challenge, however, is the small sample size compared to the large number of parameters. To address this, we will focus on optimizing the results for feature reduction in order to reduce the number of parameters for the prediction model.

References

1. Azadkia, M., Chatterjee, S., Matloff, N.: FOCI: Feature Ordering by Conditional Independence. Available at https://cran.r-project.org/package=FOCI
2. Burchett, W.W., Ellis, A.R., Harrar, S.W., Bathke, A.C.: Nonparametric Inference for Multivariate Data: The R Package npmv. J. Stat. Softw. 1–18 (2017)
3. Griessenberger, F., Junker, R.R., Trutschnig, W.: On a multivariate copula-based dependence measure and its estimation. Electron. J. Stat. **16**, 2206–2251 (2022)
4. Kuhn, M.: caret: Classification and Regression Training. Available at https://cran.r-project.org/package=caret
5. Liaw, A., Wiener, M.: Classification and Regression by random Forest. R News **2**, 18–22 (2002)
6. Neubert, S., Schreyer, M., Gerber, A., Pucher, P.: Umsetzung von Big-Data-Analysen zur Verbesserung der Walzbarrenqualität. Aluminium (2023)
7. Prillhofer, B., Antrekowitsch, H., Bottcher, H., En-Right, P.: Nonmetallic inclusions in the secondary aluminium industry for the production of aerospace alloys. In: Light Metals, pp. 603–608 (2008)
8. R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022)

9. Rudin, W.: Real and Complex Analysis, 3nd edn. Mc Graw Hill International Edition (1987)
10. Schreyer. M., Tschimpke. M., Neubert. S., Gerber. A., Trutschnig. W.: Statistical methods for measuring and predicting the quality of aluminum slabs. Light Metal Age Mag. 46–53 (2023)
11. Tschimpke, M., Trutschnig, W., Schreyer, M., Neubert, S., Gerber, A.: Statistical methods for prediction of the quality of aluminium ingots. Key Eng. Mater. **968**, 1210–1217 (2023). https://doi.org/10.4028/p-MO8olC

Fuzzy Medoids Based on an L^1-Type Distance

Beatriz Sinova[1(✉)], Sergio Palacio Vega[2], and María Ángeles Gil[1,3]

[1] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
sinovabeatriz@uniovi.es
[2] McKinsey & Company, Madrid, Spain
[3] Spanish Royal Academy of Sciences, C/ Valverde, 22, 28004 Madrid, Spain

Abstract. The first robust central tendency measures for fuzzy number-valued data introduced in the literature were extensions of the notion of median in the real-valued settings. In particular, they were defined as the fuzzy numbers that minimize the mean distance to the sample fuzzy number-valued observations. In order to solve the minimization problem, only L^1-type metrics were considered in the fuzzy-valued settings, and the corresponding measures were proven to present some interesting properties. However, these approaches do not necessarily keep the shape of the sample data, and they do not have to coincide with a sample observation like it happens in real-valued scenarios. The aim is to propose new location measures (fuzzy medoids) by restricting the previous minimization problem to the set of fuzzy observations. Their practical interest is highlighted by means of a real-life example.

Keywords: Fuzzy-valued data · Median · Medoid · Robustness

1 Preliminaries

Fuzzy-valued data are useful for imprecision modelling, which has made their interest grow in many fields, such as Psychology, Engineering or Biomedicine. For example, random experiments related to human valuations, perceptions or opinions are essentially imprecise. This work focuses on one-dimensional fuzzy sets.

Definition 1. A (bounded) **fuzzy number** is a mapping $\widetilde{U} : \mathbb{R} \to [0,1]$ such that its α-levels

$$\widetilde{U}_\alpha = \begin{cases} \{x \in \mathbb{R} : \widetilde{U}(x) \geq \alpha\} & \text{if } \alpha \in (0,1], \\ \text{cl}\{x \in \mathbb{R} : \widetilde{U}(x) > 0\} & \text{if } \alpha = 0, \end{cases}$$

where cl denotes the closure, are nonempty compact convex sets.

Let $\mathcal{F}_c(\mathbb{R})$ denote the space of all fuzzy numbers. There exist some representations that uniquely characterize any fuzzy number $\widetilde{U}$. For instance,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 466–473, 2024.
https://doi.org/10.1007/978-3-031-65993-5_57

- the infima and suprema of all α-levels of the fuzzy number,
- the midpoint/centre of a fixed (but arbitrary) α-level, $\operatorname{mid}\widetilde{U}_\beta$, and the left and right deviations with repect to the previous point:

$$\operatorname{ldev}^\beta\widetilde{U}_\alpha = \operatorname{mid}\widetilde{U}_\beta - \inf\widetilde{U}_\alpha, \quad \operatorname{rdev}^\beta\widetilde{U}_\alpha = \sup\widetilde{U}_\alpha - \operatorname{mid}\widetilde{U}_\beta.$$

The usual arithmetic can be expressed in terms of such characterizations. Given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c(\mathbb{R})$ and $\gamma \in \mathbb{R}$, the sum and product defined by Zadeh's extension principle are equivalent to the level-wise generalization of the interval arithmetic:

$$(\widetilde{U} + \widetilde{V})_\alpha = \widetilde{U}_\alpha + \widetilde{V}_\alpha, \quad (\gamma \cdot \widetilde{U})_\alpha = \gamma \cdot \widetilde{U}_\alpha.$$

Therefore, $\inf(\widetilde{U}+\widetilde{V})_\alpha = \inf\widetilde{U}_\alpha + \inf\widetilde{V}_\alpha$ and $\sup(\widetilde{U}+\widetilde{V})_\alpha = \sup\widetilde{U}_\alpha + \sup\widetilde{V}_\alpha$. If γ is nonnegative, $\inf(\gamma\cdot\widetilde{U})_\alpha = \gamma\inf\widetilde{U}_\alpha$ and $\sup(\gamma\cdot\widetilde{U})_\alpha = \gamma\sup\widetilde{U}_\alpha$; otherwise $\inf(\gamma\cdot\widetilde{U})_\alpha = \gamma\sup\widetilde{U}_\alpha$ and $\sup(\gamma\cdot\widetilde{U})_\alpha = \gamma\inf\widetilde{U}_\alpha$. Alternatively, for the sum we get

$$\operatorname{mid}(\widetilde{U} + \widetilde{V})_\beta = \operatorname{mid}\widetilde{U}_\beta + \operatorname{mid}\widetilde{V}_\beta,$$

$$\operatorname{ldev}^\beta(\widetilde{U} + \widetilde{V})_\alpha = \operatorname{ldev}^\beta\widetilde{U}_\alpha + \operatorname{ldev}^\beta\widetilde{V}_\alpha, \quad \operatorname{rdev}^\beta(\widetilde{U} + \widetilde{V})_\alpha = \operatorname{rdev}^\beta\widetilde{U}_\alpha + \operatorname{rdev}^\beta\widetilde{V}_\alpha,$$

whereas for the product by scalars it holds that $\operatorname{mid}(\gamma\cdot\widetilde{U})_\beta = \gamma\operatorname{mid}\widetilde{U}_\beta$ and

$$\operatorname{ldev}^\beta(\gamma\cdot\widetilde{U})_\alpha = \begin{cases} \gamma\operatorname{ldev}^\beta\widetilde{U}_\alpha & \text{if } \gamma \geq 0, \\ -\gamma\operatorname{rdev}^\beta\widetilde{U}_\alpha & \text{if } \gamma < 0, \end{cases}$$

$$\operatorname{rdev}^\beta(\gamma\cdot\widetilde{U})_\alpha = \begin{cases} \gamma\operatorname{rdev}^\beta\widetilde{U}_\alpha & \text{if } \gamma \geq 0, \\ -\gamma\operatorname{ldev}^\beta\widetilde{U}_\alpha & \text{if } \gamma < 0. \end{cases}$$

Among the range of statistical techniques already available for analyzing fuzzy-valued data, we can find several central tendency measures. The best-known measure is the Aumann-type mean, but, as a generalization of the mean of a real-valued random variable, it lacks robustness. The relevance of a robust behaviour when coping with outliers or data changes/errors becomes obvious when we are dealing with data collected from real-life experiments, which usually contain some 'contaminated observations'.

Definition 2. Given a sample of fuzzy number-valued observations, $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, the Aumann-type mean is the fuzzy number $\overline{\widetilde{x}}$ with α-levels given by

$$\left(\overline{\widetilde{x}}\right)_\alpha = \left[\frac{1}{n}\sum_{i=1}^n \inf(\widetilde{x}_i)_\alpha, \frac{1}{n}\sum_{i=1}^n \sup(\widetilde{x}_i)_\alpha\right]$$

for any $\alpha \in [0,1]$.

1.1 Fuzzy Medians

The first proposal in the literature was to adapt the notion of median to the fuzzy-valued case. Instead of defining the fuzzy median as a central position value with respect to certain order in $\mathcal{F}_c(\mathbb{R})$, which would be controversial due to the lack of a total order in $\mathcal{F}_c(\mathbb{R})$ that is universally accepted, it was introduced as the fuzzy numbers that minimize the mean distance to the sample fuzzy number-valued observations [3].

Definition 3. Given a sample of fuzzy number-valued observations, $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, the sample 1-norm median is any fuzzy number $\widetilde{U} \in \mathcal{F}_c(\mathbb{R})$ that minimizes

$$\frac{1}{n}\sum_{i=1}^{n} \rho_1(\widetilde{x}_i, \widetilde{U}), \tag{1}$$

where ρ_1 denotes the 1-norm distance between two fuzzy numbers, that is,

$$\rho_1(\widetilde{U}, \widetilde{V}) = \frac{1}{2}\int_{[0,1]} |\inf \widetilde{U}_\alpha - \inf \widetilde{V}_\alpha| \, d\alpha + \frac{1}{2}\int_{[0,1]} |\sup \widetilde{U}_\alpha - \sup \widetilde{V}_\alpha| \, d\alpha.$$

A particular solution to the minimization problem (1) is provided by the fuzzy number whose α-levels coincide with

$$\left(\widetilde{\mathrm{Me}}(\widetilde{x}_1, \ldots, \widetilde{x}_n)\right)_\alpha = [\mathrm{Me}\{\inf(\widetilde{x}_1)_\alpha, \ldots, \inf(\widetilde{x}_n)_\alpha\}, \mathrm{Me}\{\sup(\widetilde{x}_1)_\alpha, \ldots, \sup(\widetilde{x}_n)_\alpha\}]$$

for any $\alpha \in [0,1]$. In case any of these medians is not unique, one can consider the usual convention of choosing the midpoint of the interval of possible medians. From now on, the sample 1-norm median will refer to this specific solution $\widetilde{\mathrm{Me}}$.

Naturally, other choices of the distance among fuzzy numbers could be considered. In general, the use of L^1-type metrics is preferred since they make the study of the minimization problem easier (for the choice of an L^2-type distance in the particular case of interval-valued data, see [4]). A second approach in the literature uses an L^1-type distance based on the $\mathrm{mid}_\beta/\mathrm{ldev}^\beta/\mathrm{rdev}^\beta$ characterization of fuzzy numbers [2].

Definition 4. Given a sample of fuzzy number-valued observations, $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $\beta \in [0,1]$ and $\theta > 0$, the sample β-median is any fuzzy number $\widetilde{U} \in \mathcal{F}_c(\mathbb{R})$ that minimizes

$$\frac{1}{n}\sum_{i=1}^{n} \mathfrak{D}_\theta^\beta(\widetilde{x}_i, \widetilde{U}), \tag{2}$$

where $\mathfrak{D}_\theta^\beta(\widetilde{U}, \widetilde{V})$ denotes the following distance

$$|\mathrm{mid}\, \widetilde{U}_\beta - \mathrm{mid}\, \widetilde{V}_\beta| + \frac{\theta}{2}\int_{[0,1]} |\mathrm{ldev}^\beta \widetilde{U}_\alpha - \mathrm{ldev}^\beta \widetilde{V}_\alpha| \, d\alpha + \frac{\theta}{2}\int_{[0,1]} |\mathrm{rdev}^\beta \widetilde{U}_\alpha - \mathrm{rdev}^\beta \widetilde{V}_\alpha| \, d\alpha.$$

The nonnegative parameter θ plays the role of weighing the relative importance of the deviation in shape/imprecision with respect to the deviation in location.

It has been proven that the fuzzy number $\widetilde{\mathrm{Me}}^{\beta}$ given by the α-levels

$$\Big[\mathrm{Me}\{\mathrm{mid}\,(\widetilde{x}_1)_\beta,\ldots,\mathrm{mid}\,(\widetilde{x}_n)_\beta\}-\mathrm{Me}\{\mathrm{ldev}^\beta(\widetilde{x}_1)_\alpha,\ldots,\mathrm{ldev}^\beta(\widetilde{x}_n)_\alpha\},$$

$$\mathrm{Me}\{\mathrm{mid}\,(\widetilde{x}_1)_\beta,\ldots,\mathrm{mid}\,(\widetilde{x}_n)_\beta\}+\mathrm{Me}\{\mathrm{rdev}^\beta(\widetilde{x}_1)_\alpha,\ldots,\mathrm{rdev}^\beta(\widetilde{x}_n)_\alpha\}\Big],$$

for any $\alpha\in[0,1]$, is a solution of (2). The convention when there is a lack of uniqueness can also be considered. From now on, the sample β-median will refer to $\widetilde{\mathrm{Me}}^{\beta}$.

Proposition 1 shows that the sample 1-norm and β-medians fulfill similar properties.

Proposition 1. *For any choice of $\beta\in[0,1]$,*

- $\widetilde{\mathrm{Me}}$ *and* $\widetilde{\mathrm{Me}}^{\beta}$ *are 'affine' equivariant: if $\gamma\in\mathbb{R}$, $\widetilde{U}\in\mathcal{F}_c(\mathbb{R})$ and $\widetilde{\mathbf{x}}_n=(\widetilde{x}_1,\ldots,\widetilde{x}_n)$ is a sample of fuzzy number-valued observations, $\widetilde{\mathrm{Me}}(\gamma\cdot\widetilde{\mathbf{x}}_n+\widetilde{U})=\gamma\cdot\widetilde{\mathrm{Me}}(\widetilde{\mathbf{x}}_n)+\widetilde{U}$ and $\widetilde{\mathrm{Me}}^{\beta}(\gamma\cdot\widetilde{\mathbf{x}}_n+\widetilde{U})=\gamma\cdot\widetilde{\mathrm{Me}}^{\beta}(\widetilde{\mathbf{x}}_n)+\widetilde{U}$.*
  Then, if all the fuzzy number-valued observations coincide with $\widetilde{U}\in\mathcal{F}_c(\mathbb{R})$, both fuzzy medians are equal to $\widetilde{U}$.
- $\widetilde{\mathrm{Me}}$ *and* $\widetilde{\mathrm{Me}}^{\beta}$ *are robust: their finite sample breakdown point (minimum proportion of sample observations to be modified in order to arbitrarily increase the distance between the original fuzzy median and the fuzzy median of the modified data) is the highest for location measures.*

More properties could be stated, such as the strong consistency, if the corresponding population fuzzy medians had been introduced (see [1–3] for details). Nevertheless, the sample 1-norm and β-medians do not have to coincide with a sample observation, and this is an important difference with respect to what happens in real-valued scenarios.

Example 1. In this example, we consider a sample of three fuzzy numbers with a trapezoidal shape (see Fig. 1 left). Trapezoidal fuzzy numbers are very common in real-life applications due to simplicity reasons: it is enough to select two intervals (the 0-level and the 1-level, with a clear interpretation) and interpolate the corresponding extremes linearly to build each fuzzy number, instead of having to decide the whole shape of a general fuzzy number. Clearly, this also has important implications on the computation. Figure 1 shows both the sample 1-norm median (centre) and the β-median (right) when β was chosen to be 0.5.

The fuzzy medians do not coincide with any of the three sample observations. Apart from this, notice that neither of the fuzzy medians preserve the shape of the sample data. Even when all observations have a trapezoidal shape, these medians do not need to be a trapezoidal fuzzy number.

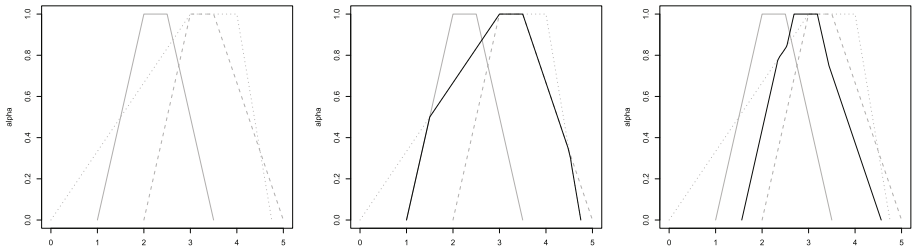


Fig. 1. Sample of 3 trapezoidal fuzzy numbers (left), their sample 1-norm median (centre, in black) and their sample 0.5-median (right, in black)

2 Fuzzy Medoids

The proposal is to constrain the minimization problems (1) and (2) to the set of fuzzy-valued observations to guarantee that the new location measures both coincide with a sample observation and keep the shape of the observations.

Definition 5. Given a sample of fuzzy number-valued observations, $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, the 1-norm medoid is any fuzzy number $\widetilde{\text{Med}}$ that minimizes the expression

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(\widetilde{x}_i, \widetilde{U}) \tag{3}$$

over $\widetilde{U} \in \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}$.

Definition 6. Given a sample of fuzzy number-valued observations, $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $\beta \in [0,1]$ and $\theta > 0$, the β-medoid is any fuzzy number $\widetilde{\text{Med}}_\theta^{\,\beta}$ that minimizes, over $\widetilde{U} \in \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathfrak{D}_\theta^\beta(\widetilde{x}_i, \widetilde{U}). \tag{4}$$

It is obvious from (3) and (4) that fuzzy medoids always exist due to the constraint that they belong to a finite subset of $\mathcal{F}_c(\mathbb{R})$. However, they could not be unique, as the following example shows.

Example 2. Fig. 2 (left) presents a sample of four trapezoidal fuzzy numbers. In this example, the solutions to the minimization problems (3) and (4) with $\beta = 0.5$ and $\theta = 1/3$ are exactly the same and not unique.

The following theoretical result proves that the introduced fuzzy medoids are also equivariant under linear transformations.

Proposition 2. *$\widetilde{\text{Med}}$ and $\widetilde{\text{Med}}_\theta^{\,\beta}$, for any choice of $\beta \in [0,1]$ and $\theta > 0$, are 'affine' equivariant: if $\gamma \in \mathbb{R}$, $\widetilde{U} \in \mathcal{F}_c(\mathbb{R})$ and $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is a sample of fuzzy number-valued observations, $\widetilde{\text{Med}}(\gamma \cdot \widetilde{\mathbf{x}}_n + \widetilde{U}) = \gamma \cdot \widetilde{\text{Med}}(\widetilde{\mathbf{x}}_n) + \widetilde{U}$ and $\widetilde{\text{Med}}_\theta^{\,\beta}(\gamma \cdot \widetilde{\mathbf{x}}_n + \widetilde{U}) = \gamma \cdot \widetilde{\text{Med}}_\theta^{\,\beta}(\widetilde{\mathbf{x}}_n) + \widetilde{U}$.*

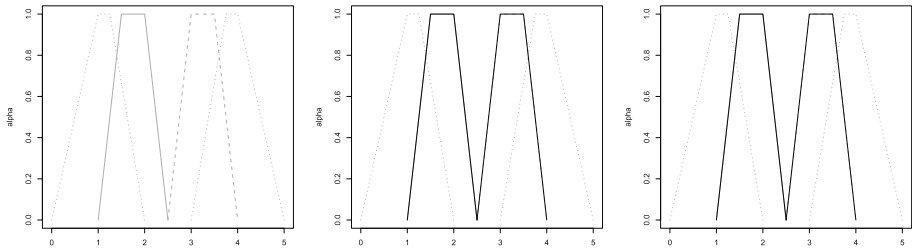


Fig. 2. Sample of 4 trapezoidal fuzzy numbers (left), their 1-norm medoids (centre, in black) and their 0.5-medoids (right, in black)

Proof. By definition of fuzzy 1-norm medoid,

$$\widetilde{\mathrm{Med}}(\gamma \cdot \widetilde{\mathbf{x}}_n + \widetilde{U}) = \arg \min_{\widetilde{V} \in \{\gamma \cdot \widetilde{x}_1 + \widetilde{U}, \ldots, \gamma \cdot \widetilde{x}_n + \widetilde{U}\}} \frac{1}{n} \sum_{i=1}^{n} \rho_1(\gamma \cdot \widetilde{x}_i + \widetilde{U}, \widetilde{V}),$$

or, alternatively,

$$\widetilde{\mathrm{Med}}(\gamma \cdot \widetilde{\mathbf{x}}_n + \widetilde{U}) = \gamma \cdot \left(\arg \min_{\widetilde{V}^* \in \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}} \frac{1}{n} \sum_{i=1}^{n} \rho_1(\gamma \cdot \widetilde{x}_i + \widetilde{U}, \gamma \cdot \widetilde{V}^* + \widetilde{U}) \right) + \widetilde{U}.$$

Taking into account that ρ_1 is translation invariant and scale equivariant,

$$\widetilde{\mathrm{Med}}(\gamma \cdot \widetilde{\mathbf{x}}_n + \widetilde{U}) = \gamma \cdot \left(\arg \min_{\widetilde{V}^* \in \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}} \frac{1}{n} \sum_{i=1}^{n} |\gamma| \rho_1(\widetilde{x}_i, \widetilde{V}^*) \right) + \widetilde{U}$$

$$= \gamma \cdot \left(\arg \min_{\widetilde{V}^* \in \{\widetilde{x}_1, \ldots, \widetilde{x}_n\}} \frac{1}{n} \sum_{i=1}^{n} \rho_1(\widetilde{x}_i, \widetilde{V}^*) \right) + \widetilde{U} = \gamma \cdot \widetilde{\mathrm{Med}}(\widetilde{\mathbf{x}}_n) + \widetilde{U}.$$

The proof for the β-medoid is analogous, since it also fulfills the translation invariance and the scale equivariance. □

If all the fuzzy number-valued observations coincide with $\widetilde{U} \in \mathcal{F}_c(\mathbb{R})$, clearly the 1-norm and β-medoids coincide with $\widetilde{U}$ too.

As a very preliminary thought about the robustness of the fuzzy medoids, the next section compares the values of the previous central tendency measures with and without contamination by means of a real-life example.

3 Real-Life Example

In 2014, a survey was carried out at Colegio San Ignacio (Oviedo, Spain). The questionnaire was based on the well-known TIMSS/PIRLS (Trends in International Mathematics and Science Study-Progress in International Reading Literacy Study) joint study in 2011, which assessed the 4th grade students in Mathematics, Science and Reading. TIMSS/PIRLS questionnaire mainly used a 4-point

Likert scale, and participants had to choose their response among the following four options: "disagree a lot", "disagree a little", "agree a little" and "agree a lot". In order to allow richer and more accurate and individually different responses, the survey used some adapted questions from the TIMSS/PIRLSS questionnaire and required a double answer in terms of both the 4-point Likert scale and a fuzzy rating scale (with trapezoidal fuzzy numbers). The survey was conducted on 69 fourth grade students.

Figure 3 (top left) shows the answers provided by the 38 male students about the item "My teacher of Mathematics is easy to understand" (one of them is not available because the student did not answer such an item). There exists an atypical datum, which corresponds to the student who selected the smallest values (the sample observation that lies "more on the left").

The Aumann-type mean, the fuzzy medians and the fuzzy medoids have been computed for the whole sample and the reduced sample, after the removal of the outlying observation. The values of the parameters have been chosen to be $\beta = 0.5$ and $\theta = 1/3$.

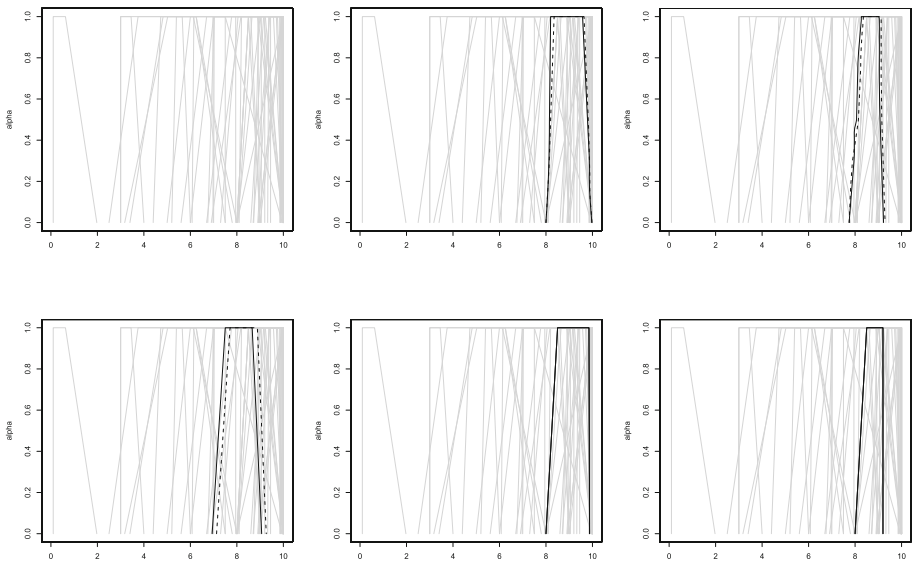


Fig. 3. Sample of the 38 responses by male students to the item "My teacher of Mathematics is easy to understand"(top left). In black, the sample Aumann-type mean (bottom left), 1-norm median (top centre), 0.5-median (top right), 1-norm medoid (bottom centre) and 0.5-medoid (bottom right) for all the observations (—) and after the removal of the atypical observation "on the left" (- -)

Results. As expected, Fig. 3 shows that the Aumann-type mean is the most sensitive central tendency measure to the outlier removal, among those considered. It is very interesting to notice that neither of the fuzzy medoids have experienced any impact under contamination, since their values for the whole and the

reduced samples exactly coincide, while the central tendency information they capture is still very similar to that of fuzzy medians.

4 Final Conclusions

The proposed fuzzy medoids present several important advantages with respect to the corresponding fuzzy medians. Naturally, when the optimization problem that defines them is constrained, the new measures do not improve, in that way, the previous results, but they can provide very competitive alternatives. Apart from preserving some convenient properties, such as the "affine" equivariance and, apparently, the robustness, it is guaranteed that they coincide with certain sample observations and, therefore, they also keep the data shape. Trapezoidal and triangular fuzzy numbers are often considered for real-life applications, and, in contrast to fuzzy medians, the 1-norm and β-medoids would summarize the central tendency of such data sets by means of a trapezoidal or triangular fuzzy number, respectively. Apart from its implications to ease the interpretation for nonexpert users, this fact simplifies the computation.

Among the future research lines, we could mention the theoretical proof of the robustness of fuzzy medoids, the analysis of more properties and the development of an empirical comparative study with the fuzzy medians and other central tendency measures for fuzzy number-valued data.

Acknowledgements. The research by Beatriz Sinova and María Ángeles Gil was partially supported by the Spanish Ministry of Science and Innovation (Grant MTM-PID2019-104486GB-I00), the Principality of Asturias (Grant SV-PA-21-AYUD/2021/50897) and INGENIERÍA DE PRESAS S.L. (Contract no. FUO-24-066), whereas Sergio Palacio Vega was granted with a Beca de Colaboración from the Spanish Ministry of Education, Vocational Training and Sports during his research collaboration. Their financial support is gratefully acknowledged.

References

1. Sinova, B., Casals, M.R., Gil, M.Á.: Central tendency for symmetric random fuzzy numbers. Inf. Sci. **278**, 599–613 (2014)
2. Sinova, B., de la Rosa de Sáa, S., Gil, M.Á.: A generalized L1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data. Inf. Sci. **242**, 22–34 (2013)
3. Sinova, B., Gil, M.Á., Colubi, A., Van Aelst, S.: The median of a random fuzzy number. The 1-norm distance approach. Fuzzy Sets Syst. **200**, 99–115 (2012)
4. Sinova, B., Van Aelst, S.: A spatial-type interval-valued median for random intervals. Statistics **52**(3), 479–502 (2018)

Fuzzy S-Estimators

Beatriz Sinova[1] and Stefan Van Aelst[2(✉)]

[1] Departamento de Estadistica e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
sinovabeatriz@uniovi.es

[2] Department of Mathematics, Section of Statistics and Data Science, KU Leuven, Leuven, Belgium
stefan.vanaelst@kuleuven.be

Abstract. The central tendency and scale of the distribution of a random fuzzy number have been estimated through several measures in the literature. The best-known measures, the Aumann-type mean and the standard deviation, lack robustness and become unreliable under data contamination. With respect to the other approaches, with a more robust behaviour, scale estimators corresponding to the fuzzy medians and fuzzy trimmed means are available, but there is not any natural choice regarding M-estimators of location. In the real-valued setting, S-estimators of location present a similar behaviour to M-estimators and they are associated with a scale estimator. Therefore, the notion of S-estimator will be adapted to the fuzzy-valued case and its practical interest will be illustrated by means of some simulation studies.

Keywords: Fuzzy number · Location · S-estimator · Simulation · Robustness

1 Introduction

Fuzzy-valued data are well suited to model intrinsically imprecise attributes. It is challenging to reliably analyse this type of data. For the statistical analysis of fuzzy-valued data it is often assumed that they are measurements of a random fuzzy number (or fuzzy random variable in Puri and Ralescu's sense [9]). A main interest is to estimate the location (central tendency) and scale (dispersion) of the distribution of such a random fuzzy number. Popular location and scale measures in this setting are the Aumann-type mean [9] and the variance or standard deviation of a random fuzzy number [7,8]. While these measures satisfy many useful properties, as in the real-valued setting they are also highly sensitive to anomalies in the data such as outliers or small changes in the data. Therefore, there is a need for more robust measures of location and scale that are less affected by such anomalies.

Recent proposals to measure the location of a random fuzzy number include the 1-norm median [13], the wabl/ldev/rdev median [12], trimmed means [3]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 474–482, 2024.
https://doi.org/10.1007/978-3-031-65993-5_58

and M-estimators of location [14]. Scale estimators corresponding to median-type location measures have been developed in [5]. M-estimators can be recommended for robust estimation of location because of their good performance, both in terms of accuracy and robustness [16,17]. However, in contrast to other robust estimators such as trimmed means, a corresponding scale estimator is not naturally available for location M-estimators. In the real setting, S-estimators have been introduced to remedy this [10], since the S-estimator of location behaves similarly as an M-estimator of location, while it comes with an associated S-estimator of scale. Therefore, the main goal of this paper is to introduce the adaptation of S-estimators to the fuzzy-valued setting.

2 Preliminary Concepts and Notations

Fuzzy-valued data are a natural way to describe imprecise human assessments such as ratings, opinions, judgments or perceptions [4]. Formally, the space of fuzzy numbers can be defined as follows.

Definition 1. $\mathcal{F}_c^*(\mathbb{R})$ is the space of (bounded) **fuzzy numbers**, that is, the mappings $\widetilde{U} : \mathbb{R} \to [0,1]$ such that their α-levels

$$\widetilde{U}_\alpha = \begin{cases} \{x \in \mathbb{R} : \widetilde{U}(x) \geq \alpha\} & \text{if } \alpha \in (0,1] \\ \mathrm{cl}\{x \in \mathbb{R} : \widetilde{U}(x) > 0\} & \text{if } \alpha = 0, \end{cases}$$

are nonempty compact intervals. $\widetilde{U}(x)$ can be interpreted as the 'degree of compatibility of x with $\widetilde{U}$' (or 'degree of truth' of the assertion "x is $\widetilde{U}$").

The most common approach to define an arithmetic for the elements of this space is based on Zadeh's extension principle, which is a level-wise extension of the usual interval arithmetic.

Definition 2. Let $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$. The **sum** of $\widetilde{U}$ and $\widetilde{V}$ is defined as the fuzzy number $\widetilde{U} + \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ given for each $\alpha \in [0,1]$ by

$$\begin{aligned} (\widetilde{U} + \widetilde{V})_\alpha &= \text{Minkowski sum of } \widetilde{U}_\alpha \text{ and } \widetilde{V}_\alpha \\ &= \left[\inf \widetilde{U}_\alpha + \inf \widetilde{V}_\alpha, \sup \widetilde{U}_\alpha + \sup \widetilde{V}_\alpha\right]. \end{aligned}$$

Let $\widetilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ and $\gamma \in \mathbb{R}$. The **product** of $\widetilde{U}$ **by the scalar** γ is defined as the fuzzy number $\gamma \cdot \widetilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ given for each $\alpha \in [0,1]$ by

$$(\gamma \cdot \widetilde{U})_\alpha = \gamma \cdot \widetilde{U}_\alpha = \begin{cases} \left[\gamma \cdot \inf \widetilde{U}_\alpha, \gamma \cdot \sup \widetilde{U}_\alpha\right] & \text{if } \gamma \geq 0 \\ \left[\gamma \cdot \sup \widetilde{U}_\alpha, \gamma \cdot \inf \widetilde{U}_\alpha\right] & \text{otherwise.} \end{cases}$$

It is well-known that with this arithmetic the space $\mathcal{F}_c^*(\mathbb{R})$ is a semi-linear space rather than a linear space. This implies that it is not possible to define a

difference between fuzzy values that preserves the main properties of the difference between real values in relation to the sum. Therefore, statistical methods for imprecise data make use of suitable distances between fuzzy numbers. Equality of fuzzy numbers then corresponds to the distance between these fuzzy values being equal to 0. By using appropriate distances, fuzzy number-valued data can be embedded in a functional data space through their so-called support function [9]. However, this does not mean that fuzzy-valued data can be analyzed by functional data methods in general because these methods cannot guarantee that the results are still fuzzy numbers.

In this paper, we consider the versatile parameterized family of distances introduced by [2] and defined as follows:

Definition 3. Let $\theta \in (0, +\infty)$ and let φ be an absolutely continuous probability measure on $([0,1], \mathcal{B}_{[0,1]})$ with the mass function being positive on $(0,1)$. The **mid/spr-based L^2 distance** is defined as the mapping $D_\theta^\varphi : \mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R}) \to [0, +\infty)$ which associates with each pair $\widetilde{U}, \widetilde{V}$ of elements from $\mathcal{F}_c^*(\mathbb{R})$, the value

$$D_\theta^\varphi(\widetilde{U}, \widetilde{V}) = \left[\int_{[0,1]} \left(\operatorname{mid} \widetilde{U}_\alpha - \operatorname{mid} \widetilde{V}_\alpha \right)^2 d\varphi(\alpha) + \theta \int_{[0,1]} \left(\operatorname{spr} \widetilde{U}_\alpha - \operatorname{spr} \widetilde{V}_\alpha \right)^2 d\varphi(\alpha) \right]^{1/2},$$

where $\operatorname{mid} \widetilde{U}_\alpha = (\inf \widetilde{U}_\alpha + \sup \widetilde{U}_\alpha)/2$ and $\operatorname{spr} \widetilde{U}_\alpha = (\sup \widetilde{U}_\alpha - \inf \widetilde{U}_\alpha)/2$.

This metric has been used successfully for the development of statistical methods for the analysis of fuzzy-valued data, see e.g. the review in [1].

This family of distances takes differences in center and shape into account through the mid and spr functions, respectively. The parameter θ weighs the influence of the 'deviation in shape' between the fuzzy-valued data with respect to the influence of their 'deviation in center'. Two popular choices are $\theta = 1$, which corresponds to a generalization of the well-known distance $\boldsymbol{\rho_2}$ in [6], and $\theta = 1/3$ because in this case the distance can equivalently be expressed as

$$D_{1/3}^\varphi(\widetilde{U}, \widetilde{V}) = \sqrt{\int_{[0,1]} \int_{[0,1]} \left(\widetilde{U}_\alpha^{[\eta]} - \widetilde{V}_\alpha^{[\eta]} \right)^2 d\varphi(\alpha)\, d\ell(\eta)},$$

with ℓ the Lebesgue measure on $([0,1], \mathcal{B}_{[0,1]})$ and $\widetilde{U}_\alpha^{[\eta]} = \eta \cdot \sup \widetilde{U}_\alpha + (1-\eta) \cdot \inf \widetilde{U}_\alpha$ for all $\eta \in [0,1]$.

The notion of random fuzzy number [9] is an essential tool to formalize the statistical analysis of fuzzy-valued data. It assumes that fuzziness affects the nature of values/data while randomness affects their generation.

Definition 4. Let $(\Omega, \mathcal{A}, P)$ be a probability space modelling a random experiment. A mapping $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R})$ is said to be a **random fuzzy number**

associated with the random experiment if, and only if, for each $\alpha \in [0,1]$ the interval-valued mapping $\mathcal{X}_\alpha$ (where $\mathcal{X}_\alpha(\omega) = \big(\mathcal{X}(\omega)\big)_\alpha$ for all $\omega \in \Omega$) is a random compact interval or equivalently, the real-valued functions $\inf \mathcal{X}_\alpha$ and $\sup \mathcal{X}_\alpha$ are random variables such that $\inf \mathcal{X}_\alpha(\omega) \leq \sup \mathcal{X}_\alpha(\omega)$ for all $\omega \in \Omega$.

Note that a random fuzzy number is a Borel-measurable mapping with respect to the Borel σ-field associated with several metrics, including the mid/spr-based family of distances. Consequently, crucial concepts such as the *induced distribution of a random fuzzy number* or the *stochastic independence of random fuzzy numbers* can be immediately stated.

The most popular measure of location of a random fuzzy number is the Aumann-type mean [9], which is defined as follows.

Definition 5. Let $\mathcal{X}$ be a random fuzzy number and assume that the expected values of the random variables $\inf \mathcal{X}_0$ and $\sup \mathcal{X}_0$ are finite. Then, the **Aumann-type mean** of $\mathcal{X}$ is the fuzzy number $\widetilde{E}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$ such that for each $\alpha \in [0,1]$

$$\big(\widetilde{E}(\mathcal{X})\big)_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)]\,.$$

The Aumann-type mean is also the fuzzy number minimizing the 'mean square error' of $\mathcal{X}$ about a fuzzy number. That is,

$$\widetilde{E}(\mathcal{X}) = \arg \min_{\widetilde{U} \in \mathcal{F}_c^*(\mathbb{R})} E\left[\left(D_\theta^\varphi(\mathcal{X}, \widetilde{U})\right)^2\right]$$

for every θ and φ for which the expectation exists.

A natural scale measure corresponding to the Aumann-type mean as location measure of a random fuzzy number $\mathcal{X}$ is the standard deviation which, if it exists, is the real number $\sigma(\mathcal{X})$ given by

$$\sigma(\mathcal{X}) = \sqrt{E\left[\left(D_\theta^\varphi(\mathcal{X}, \widetilde{E}(\mathcal{X})\right)^2\right]}.$$

The standard deviation $\sigma(\mathcal{X})$ corresponds to the square root of the variance of the random fuzzy number considered in [7,8].

3 S-Estimators of Location and Scale for Fuzzy-Valued Data

We now introduce location-scale S-estimators for random fuzzy numbers following the definition of S-estimators in real-valued settings [10]. To this end, we consider a loss function $\rho : [0,\infty) \to [0,\infty)$ which satisfies

(A1) ρ has a continuous derivative and $\rho(0) = 0$.
(A2) There exists a finite constant $k > 0$ such that ρ is strictly increasing on $[0,k]$ and constant on $[k,\infty)$.

Definition 6. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R})$ an associated random fuzzy number. Moreover, let ρ be a loss function which satisfies conditions (A1) and (A2). Then, the **fuzzy S-functional of location and scale** is the solution $(\widetilde{T}_S(\mathcal{X}), \sigma_S(\mathcal{X}))$ which minimizes σ subject to the constraint

$$E\left[\rho\left(\frac{D_\theta^\varphi(\mathcal{X}, \widetilde{T})}{\sigma}\right)\right] = b$$

over all $(\widetilde{T}, \sigma) \in \mathcal{F}_c^*(\mathbb{R}) \times (0, \infty)$, where $0 < b < \rho(k)$.

Similarly, for a simple random sample $\boldsymbol{\mathcal{X}}_n = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ from $\mathcal{X}$, the **sample fuzzy S-estimator of location and scale** is the solution $(\widehat{\widetilde{T}_S}(\boldsymbol{\mathcal{X}}_n), \widehat{\sigma_S}(\boldsymbol{\mathcal{X}}_n))$ which minimizes σ subject to the constraint

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{D_\theta^\varphi(\mathcal{X}_i, \widetilde{T})}{\sigma}\right) = b$$

over all $(\widetilde{T}, \sigma) \in \mathcal{F}_c^*(\mathbb{R}) \times (0, \infty)$, where $0 < b < \rho(k)$.

A popular choice for the loss function in the definition of S-estimators is the Tukey bisquare loss function given by

$$\rho_d^T(x) = \begin{cases} d^2/6 \cdot (1 - (1 - (x/d)^2)^3) & \text{if } 0 \le x \le d, \\ d^2/6 & \text{otherwise,} \end{cases}$$

with tuning parameter $d > 0$. This loss function has already shown good performance when estimating the location of random fuzzy numbers [15].

4 Empirical Evaluation of Fuzzy S-Estimators

To investigate the performance of fuzzy S-estimators of location and scale, we compare it to the Aumann-type mean and associated standard deviation and to the fuzzy trimmed mean and associated trimmed scale in a simulation study. We use trapezoidal-valued random fuzzy numbers, characterized by four real-valued random variables as follows: $\mathcal{X} = (X_1 - X_2 - X_3, X_1 - X_2, X_1 + X_2, X_1 - X_2 + X_4)$ with $X_1 = \text{mid}\mathcal{X}_1$, $X_2 = \text{spr}\mathcal{X}_1$, $X_3 = \inf \mathcal{X}_1 - \inf \mathcal{X}_0$ and $X_4 = \sup \mathcal{X}_0 - \sup \mathcal{X}_1$. A sample of size $n = 1000$ is generated from $\mathcal{X}$. Two cases are considered:

- CASE 1 assumes that all random variables X_i are independent and
  - $X_1 \sim \mathcal{N}(0, 1)$;
  - $X_2, X_3, X_4 \sim \chi_1^2$.
- CASE 2 assumes that there is dependence among the random variables X_i, which are modelled by the probability distributions:
  - $X_1 \sim \mathcal{N}(0, 1)$;
  - $X_2, X_3, X_4 \sim 1/(X_1^2 + 1)^2 + \sqrt{\chi_1^2}$, with χ_1^2 being independent of X_1.

Afterwards, a proportion $c_p \in \{0.1, 0.2, 0.4\}$ of sample data is contaminated.

- In CASE 1, nc_p original data are replaced by trapezoidal fuzzy numbers generated from:
  - $X_1 \sim \mathcal{N}(0,3) + C_D$;
  - $X_2, X_3, X_4 \sim \chi^2_4 + C_D$.
- In CASE 2, the contaminated data are generated from:
  - $X_1 \sim \mathcal{N}(0,3) + C_D$;
  - $X_2, X_3, X_4 \sim 1/(X_1^2+1)^2 + \sqrt{\chi^2_1} + C_D$, with χ^2_1 being independent of X_1.

Clearly the role of the parameter $C_D \in \{0, 1, 5, 10, 100\}$ is to quantify the separation between the distribution of the two subsamples, consisting of noncontaminated and contaminated observations, respectively. For each case and each contamination situation, determined by the values of the parameters c_p and C_D, we have computed the mentioned estimators of central tendency (the Aumann-type mean, the fuzzy trimmed means with trimming proportion 0.2 and 0.4 and the fuzzy S-estimator of location) and dispersion (standard deviation, the trimmed scale associated with the fuzzy trimmed means and the fuzzy S-estimator of scale) for the whole sample. In particular, the fuzzy S-estimator has been obtained by adapting the fast algorithm presented in [11] for the classical framework and considering the Tukey bisquare loss function with $d = 1$.

The mean $D^{\ell}_{1/3}$-distance between each central tendency measure and the noncontaminated data has been computed as well. These simulation steps have been repeated $N = 1000$ times and Table 1 shows for each case the average of the mean $D^{\ell}_{1/3}$-distance between the location measures and the data in the noncontaminated subsample, as an approximation of the expected $D^{\ell}_{1/3}$-distance between those measures and the noncontaminated distribution. Furthermore, Table 1 includes the average of each scale estimate for the N mixed (partially contaminated and noncontaminated) samples.

Results. The behaviour of all the measures is very similar in both cases. As expected, the Aumann-type mean and the standard deviation are highly sensitive to (even little) contamination. The performance of the fuzzy trimmed mean completely depends on the considered trimming proportion: 0.2 is not enough for dealing with $c_p = 0.4$, whereas the choice 0.4 can cope with all the values of c_p in this table. Thus, the 0.4-fuzzy trimmed mean and the standard deviation of the trimming region summarize the central tendency and the scale, respectively, in a robust way, but the results are even better for the proposed S-estimator of scale (while that of the location is similar to the 0.4-trimmed mean).

5 Final Conclusions

As explained before, the interest to analyze S-estimators of location was to be able to provide an associated scale estimator, while preserving the good behaviour of M-estimators of location. In this sense, the adaptation of fuzzy S-estimators of location and scale to the fuzzy number-valued setting has shown very competitive empirical results. Future research lines will cope with the complete study of the theoretical properties satisfied by fuzzy S-estimators.

Table 1. Mean distance of the sample Aumann-type mean (Mean), fuzzy trimmed mean with trimming proportion 0.2 (0.2-Trim) or 0.4 (0.4-Trim) and fuzzy S-estimator of location (Slocation) to the noncontaminated distribution. Monte Carlo approximation of the standard deviation (Std), the trimmed scale associated with the previous fuzzy trimmed means (0.2-STrim and 0.4-STrim, respectively) and the fuzzy S-estimator of scale (Sscale)

CASE 1									
c_p	C_D	Mean	0.2-Trim	0.4-Trim	Slocation	Std	0.2-STrim	0.4-STrim	Sscale
0	0	1.23547	1.22006	1.22616	1.24312	1.45099	0.98745	0.75523	1.23692
0.1	0	1.29854	1.22110	1.22312	1.23936	1.96688	1.10652	0.82380	1.33163
0.1	1	1.33163	1.22092	1.22325	1.23957	2.10885	1.11808	0.82629	1.33304
0.1	5	1.55257	1.22071	1.22329	1.23958	3.15909	1.12170	0.82643	1.33305
0.1	10	1.97792	1.22070	1.22329	1.23958	4.90681	1.12170	0.82643	1.33305
0.1	100	13.5516	1.22070	1.22329	1.23958	40.5687	1.12170	0.82643	1.33305
0.2	0	1.40216	1.23157	1.22083	1.23520	2.34722	1.27529	0.91282	1.44952
0.2	1	1.49769	1.23396	1.22097	1.23575	2.55211	1.32861	0.92199	1.45394
0.2	5	2.14754	1.23542	1.22105	1.23578	4.04236	1.43980	0.92304	1.45403
0.2	10	3.27043	1.23570	1.22105	1.23578	6.43367	1.45196	0.92304	1.45403
0.2	100	27.0413	1.23570	1.22105	1.23578	54.0819	1.45209	0.92304	1.45403
0.4	0	1.69401	1.36253	1.23305	1.22479	2.88545	1.75470	1.19471	1.80200
0.4	1	1.97219	1.46221	1.23584	1.22650	3.14528	1.97717	1.28128	1.83042
0.4	5	3.66405	2.17236	1.23484	1.22684	4.95895	3.49098	1.43695	1.83233
0.4	10	6.19219	3.54815	1.23514	1.22684	7.88836	6.04716	1.45154	1.83233
0.4	100	54.0485	33.2730	1.23514	1.22684	66.2334	57.6299	1.45169	1.83233
CASE 2									
c_p	C_D	Mean	0.2-Trim	0.4-Trim	Slocation	Std	0.2-STrim	0.4-STrim	Sscale
0	0	0.97699	0.97342	0.97776	0.98157	1.13372	0.81786	0.62863	1.06008
0.1	0	0.98003	0.97345	0.97646	0.98097	1.44529	0.87533	0.66386	1.11094
0.1	1	0.97724	0.97345	0.97809	0.98220	1.48485	0.89103	0.67765	1.13505
0.1	5	1.12529	0.97372	0.97571	0.98065	2.40554	0.92258	0.68720	1.14235
0.1	10	1.57549	0.97372	0.97571	0.98065	4.17558	0.92258	0.68720	1.14235
0.1	100	13.3274	0.97372	0.97571	0.98065	39.9390	0.92258	0.68720	1.14235
0.2	0	0.98408	0.97452	0.97549	0.98057	1.70004	0.94555	0.70542	1.17009
0.2	1	0.98931	0.97443	0.97987	0.98407	1.75751	0.97762	0.73527	1.22453
0.2	5	1.55528	0.97698	0.97415	0.97979	3.07527	1.13382	0.76643	1.24750
0.2	10	2.73781	0.97706	0.97415	0.97979	5.48410	1.13402	0.76643	1.24750
0.2	100	26.6205	0.97706	0.97415	0.97979	53.2452	1.13402	0.76643	1.24750
0.4	0	0.99448	0.98024	0.97456	0.97927	2.12110	1.14081	0.81279	1.31990
0.4	1	1.05057	0.98538	0.99125	0.99501	2.19156	1.20253	0.87703	1.45317
0.4	5	2.69613	1.51658	0.97732	0.97680	3.78699	2.48677	1.13400	1.57958
0.4	10	5.27960	2.95714	0.97741	0.97680	6.73078	5.07896	1.13422	1.57958
0.4	100	53.2209	32.8191	0.97741	0.97680	65.2091	56.8444	1.13422	1.57958

Acknowledgments. The research by B. Sinova was partially supported by the Spanish Ministry of Science and Innovation (Grant MTM-PID2019-104486GB-I00), the Principality of Asturias (Grant SV-PA-21-AYUD/2021/50897) and INGENIERÍA DE PRESAS S.L. (Contract no. FUO-24-066). Their financial support is gratefully acknowledged.

References

1. Blanco-Fernández, A., Casals, M.R., Colubi, A., Corral, N., García-Bárzana, M., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Sáa, S., Sinova, B.: A distance-based statistical analysis of fuzzy number-valued data. Int. J. Approx. Reason. **55**, 1487–1501 (2014)
2. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Math. Soft Comput. **2**, 71–84 (1995)
3. Colubi, A., González-Rodríguez, G.: Fuzziness in data analysis: towards accuracy and robustness. Fuzzy Sets Syst. **281**, 260–271 (2015)
4. Fuzzy rating scale-based questionnaires and their statistical analysis: de la Rosa de Sáa, S., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A. IEEE Trans. Fuzzy Syst. **23**, 111–126 (2015)
5. Robust scale estimators for fuzzy data: de la Rosa de Sáa, S., Lubiano, M.A., Sinova, B., Filzmoser, P. Adv. Data Anal. Classif. **11**, 731–758 (2017)
6. Diamond, P., Kloeden, P.: Metric spaces of fuzzy sets. Fuzzy Sets Syst. **35**, 241–249 (1990)
7. Körner, R.: On the variance of fuzzy random variables. Fuzzy Sets Syst. **92**, 83–93 (1997)
8. Lubiano, M.A., Gil, M.Á., López-Díaz, M., López-García, M.: The $\boldsymbol{\lambda}$-mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets Syst. **111**, 307–317 (2000)
9. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. **114**, 409–422 (1986)
10. Rousseeuw, P.R., Yohai, V.: Robust Regression by means of S-estimators. In: Franke, J., Haerdle, W., Martin, D. (eds.) Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics, vol. 26, pp. 256–272. Springer, Berlin (1984)
11. Salibian-Barrera, M., Yohai, V.J.: A Fast Algorithm for S-Regression Estimates. J. Comput. Graph. Stat. **15**, 414–427 (2006)
12. Sinova, B., de la Rosa de Sáa, S., Gil, M.Á.: A generalized L^1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data. Inf. Sci. **242**, 22–34 (2013)
13. Sinova, B., Gil, M.Á., Colubi, A., Van Aelst, S.: The median of a random fuzzy number, the 1-norm distance approach. Fuzzy Sets Syst. **200**, 99–115 (2012)
14. Sinova, B., Gil, M.Á., Van Aelst, S.: M-estimates of location for the robust central tendency of fuzzy data. IEEE Trans. Fuzzy Syst. **24**(4), 945–956 (2016)
15. Sinova, B., Van Aelst, S.: Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function. Int. J. Approx. Reason. **93**, 219–237 (2018)
16. Sinova, B., Van Aelst, S.: Empirical comparison of the performance of location estimates of fuzzy number-valued data. In: Destercke, S., Denoeux, T., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Advances in Intelligent Systems and Computing. Uncertainty Modelling in Data Science, vol. 832, pp. 191–199. Springer, Switzerland (2019)

17. Sinova, B., Van Aelst, S.: Empirical analysis of the maximum asymptotic bias of location estimators for fuzzy number-valued data. Int. J. Approx. Reason. **113**, 1–13 (2019)

Reversing Inhomogeneous and Imprecise Markov Chains

Damjan Škulj(✉)

Faculty of Social Sciences, University of Ljubljana,
Kardeljeva pl. 5, 1000 Ljubljana, Slovenia
damjan.skulj@fdv.uni-lj.si

Abstract. Reversal of a Markov chain means observing the same process backwards in time. A class of Markov chains whose reverses adhere to the same probabilistic laws as the original processes are called reversible Markov chains. In this article, a reversal of imprecise Markov chains is proposed. This is done by reversing a set of compatible inhomogeneous Markov chains. The reversed processes can be explicitly characterized if the interpretation of strong independence is adopted. However, the reversed processes generally lack some important properties such as the Markov property.

Keywords: Markov chain · Reversibility · Imprecise Markov chains

1 Introduction

Markov chains [8,11] are popular tools for modelling the temporal evolution of processes in which the future state of a process is described using probability distributions over a *state space*. The Markov property ensures that if a state, or more generally a distribution over the states, is known at a certain point in time, the future development of the process is independent of the past.

Sometimes it is advantageous to consider the same process in reverse time, i.e. when past and future swap roles. The process obtained by reversing time is called a *reversed process*. It is known that in the precise probability settings, a reversed Markov chain again satisfies the Markov property, but its properties may change. In particular, if a process is not *stationary*, the reversed process loses its time homogeneity, which means that the transition probabilities may vary in time.

Most attention in the literature has been devoted to processes that adhere to the same probabilistic laws when reversed. Such processes are said to be *reversible*, and in particular reversible Markov chains have been extensively studied. These have been applied in many fields, such as chemistry [12], analysis of random walks [1,6], queueing theory [7] and numerical analysis [2].

An important generalisation of Markov chains that includes uncertainty in probability distributions is the theory of imprecise Markov chains [3,9]. Depending on the interpretation, imprecise Markov chains do not satisfy time homogene-

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 483–490, 2024.
https://doi.org/10.1007/978-3-031-65993-5_59

ity or even the Markov property in the classical sense. Instead, they only fulfil these properties in the sense of sets of distributions and transition probabilities.

The aim of this paper is to investigate the reversed processes in the imprecise settings. In particular, this requires an analysis of reverse inhomogeneous Markov chains, which has received little attention in the literature so far.

The structure of this paper is as follows: Sect. 2 introduces the basics of Markov chains and their reversals. Section 3 deals with the reversal of time-inhomogeneous Markov chains. Section 4 introduces imprecise Markov chains. In Sect. 5, we focus on the reversal of imprecise Markov chains. The paper concludes with a discussion in Sect. 6.

2 Markov Chains and Their Reversals

We consider Markov chains of the form (X_n), where n is typically a time index that runs over a set of integers. The random variable X_n then takes a value $x \in \mathcal{X}$, where $\mathcal{X}$ is a finite set called *state space* and whose elements are called *states*. We will mainly deal with the chains within a finite time horizon. For convenience, we introduce the notation $X_{m:n} = (X_m, \dots, X_n)$ for some time indices m and n. In the case that $m > n$, this notation will simply denote the descending sequence $(X_m, X_{m-1}, \dots, X_n)$. This notation will apply to any type of indexed objects.

A process that satisfies the *Markov property*, i.e., $P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = P(X_{n+1} = x_{n+1} \mid X_n = x_n) = P_n(x_n, x_{n+1})$ for each $n \in \mathbb{N}$, is called a Markov chain. In addition to *time-homogeneous chains*, where the transition matrices are constant in time, i.e., $P_n = P$, we also consider time-inhomogeneous Markov chains, where the transition matrices can depend on time. A vector $q\colon \mathcal{X} \to \mathbb{R}$ with non-negative components that sum to one denotes a *probability mass function* over $\mathcal{X}$ – all probability mass functions in this paper are represented with row vectors. In the case where the distribution q^1 of X_1 over $\mathcal{X}$ is known, it is called the *initial distribution*. For an initial distribution q^1, the distribution over the states at time n is equal to $q^n = q^{n-1}P_{n-1} = q^1 P_1 \dots P_{n-1}$.

The initial distribution q^1 together with a collection of transition matrices $P_{1:n-1}$ now completely determines the distributions of the chain. It is called the *transition law*, which is denoted by $\Gamma = (q^1, P_{1:n-1})$. For a time-homogeneous Markov chain, we simply write its transition law as (q^1, P). The *reversed chain* refers to the sequence $(X_{n:1})$, which is the same process running backward in time. Most analyses of reverse processes are restricted to the case of stationary time-homogeneous chains. The class of irreducible time-homogeneous Markov chains, where the joint distributions are the same when time runs forward or backward, are called *reversible Markov chains*. A reversible chain is *stationary*, which means that all its terms X_i are distributed according to the *stationary*

distribution π corresponding to the transition matrix P, which satisfies the condition $\pi P = \pi$. The reversed transition matrix is then constant in time and equal to

$$P^*(x,y) = \frac{\pi(y)P(y,x)}{\pi(x)}, \tag{1}$$

or in terms of matrix algebra $P^* = \mathrm{diag}(\pi)^{-1}P^t\mathrm{diag}(\pi)$. The reversed chain is again irreducible with stationary distribution π. If the assumptions are relaxed and the original chain is non-stationary, the reversed process is still Markov, but it does not satisfy time homogeneity [4]. We will discuss this case in the next section on reversing inhomogeneous Markov chains.

For a reversible Markov chain, $P^*(x,y) = P(x,y)$ holds. It follows directly from (1) that for a reversible Markov chain the following *detailed balance* condition applies:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \quad \text{for all } x,y \in \mathcal{X}, \tag{2}$$

or $\mathrm{diag}(\pi)P = P^t\mathrm{diag}(\pi)$.

3 Reversing a Time Inhomogeneous Markov Chain

The formula for the reverse transition matrix (1) provides a transition matrix of the reverse process under the condition that the process is stationary, i.e. that the marginal distributions of X_n are all π, which is a stationary distribution of the constant transition matrix P. In the time-inhomogeneous case, the transition matrices P_n depend on time, and the marginal distributions q^n corresponding to X_n are generally not stationary distributions of the transition matrices. The reversal of such a chain requires a more general approach, where the reversed transitions depend not only on the forward transitions but also on the marginal distributions.

The following definition represents the basic operation of the reversal.

Definition 1. Let P be a transition matrix and q an arbitrary probability mass function on $\mathcal{X}$. Then let

$$P_q^* = \mathrm{diag}(qP)^{-1}P^t\mathrm{diag}(q) \tag{3}$$

and we denote P_q^* as a *q-reverse* of P.

Note that in the case that π is a stationary distribution for P, the π-reverse of P is exactly the usual reversed transition matrix P^*.

The following results can be verified by direct calculations.

Proposition 1. *Let P be a stochastic matrix and q a probability mass function on $\mathcal{X}$. The following applies.*

(i) P_q^* *is a stochastic matrix;*
(ii) $qPP_q^* = q$;
(iii) $(P_q^*)_{qP}^* = P$.

The qP-reverse of the q-reverse of P is therefore P again.

Definition 2. Let $\Gamma = (q^1, P_{1:n-1})$ be a transition law of an inhomogeneous Markov chain. Then we define $\Gamma^* = (q^n, P^*_{n-1:1})$, where $P^*_k = (P_k)^*_{q^k}$, i.e. $P^*_k = \mathrm{diag}(q^{k+1})^{-1} P_k^t \mathrm{diag}(q^k)$ and $q^k = q^1 P_1 \cdots P_{k-1}$ for each k, to be the *reverse transition law* to Γ.

Proposition 2. *Let $(X_{1:n})$ be an inhomogeneous Markov chain with the transition law $\Gamma = (q^1, P_{1:n-1})$. Then its time reversal is the chain $(X_{n:1})$ and its transition law is Γ^*.*

Remark 1. Even if the transition law of a chain $(X_{1:n})$ is homogeneous with a transition matrix P, the reversed law is generally inhomogeneous with transition matrices $P^*_k = P^*_{q^k}$, unless it starts from the stationary distribution π corresponding to P. In this case, the reversed transition law is homogeneous with the usual reversed transition matrix $P^* = P^*_\pi$.

4 Imprecise Markov Chains

Markov chains, whose parameters are only partially known, have been studied under the name *imprecise Markov chains* [3,9] or *Markov set chains* [5]. An *imprecise Markov chain* is a sequence (X_k) of random variables that assume values from a set of states $\mathcal{X}$, and whose transition law is given in terms of imprecise probabilities. In most models in the literature, an imprecise transition law is assumed in terms of an initial set of distributions $\mathcal{C}$ and a set of transition matrices $\mathcal{P}$. It is then assumed that the transition probability function $P(X_k, X_{k+1})$ corresponds to any matrix in $\mathcal{P}$.

Under these assumptions, however, several different interpretations are possible, leading to different models:

Precise Markov chain with unknown parameters. This interpretation is based on a simple scenario in which the transition probabilities are time-homogeneous. However, the exact values of these probabilities are not precisely known; it is assumed that they are contained in the set $\mathcal{P}$. In this framework, we consider a particular Markov chain with fixed but unknown parameters. The transition law in this case is represented by a single transition matrix P, which is a member of the set $\mathcal{P}$.

Strong independence interpretation. In this approach, each transition matrix denoted by P_k is generally unique and lies in the set $\mathcal{P}$. A crucial aspect of this interpretation is that the transition probabilities at any given time n are constant and independent of the history of the process. This construction defines a collection of time inhomogeneous Markov chains, where each chain satisfies the Markov property.

Epistemic irrelevance interpretation. This interpretation allows the transition matrices at a given time n to be influenced by the previous state history, but still belong to the set $\mathcal{P}$. In this framework, the transition probabilities depend in principle on the entire state sequence, say $x_{1:k-1}$. So we have the conditional probability mass function $P(\cdot|x_k, x_{1:k-1})$, which we assume is in the x_k-th row set $\mathcal{P}_{x_k}$ of the set of transition matrices $\mathcal{P}$.

The processes corresponding to the three interpretations above are nested in the sense that the interpretation of the precise chain with unknown parameters implies strong independence, which in turn implies epistemic irrelevance. The reverse implications are generally not valid.

5 Reversing an Imprecise Markov Chain

An imprecise Markov chain $(X_{1:n})$ with a transition law given by the initial distribution C and the transition set $\mathcal{P}$ induces a set of probability distributions on $\mathcal{X}^n$ that depends on the interpretation explained above. We will denote the distribution set by $\text{distr}(X_{1:n})$. Now let $p \in \text{distr}(X_{1:n})$. Then we denote $p^*(x_{1:n}) = p(x_{n:1})$ as the reversed distribution, and $\text{distr}(X_{1:n})^* = \{p^* \colon p \in \text{distr}(X_{1:n})\}$. Note that in the case where $n = 2$, the probability mass function is a $|\mathcal{X}| \times |\mathcal{X}|$ probability mass matrix and p^* is then its transpose.

Proposition 3. *Let* $\text{distr}(X_{1:n})$ *be an imprecise joint distribution of a chain* $(X_{1:n})$. *Then* $\text{distr}(X_{1:n})^*$ *is the joint distribution of the reversed chain* $(X_{n:1})$.

5.1 Reversed Imprecise Markov Chains Depending on the Interpretation

According to the interpretation of a precise Markov chain with unknown parameters, the transition law of a Markov chain $(X_{1:n})$ is given by an initial distribution q^1 and a fixed transition probability matrix $P \in \mathcal{P}$. There are two possibilities depending on the initial distribution q^1. The simplest approach is to consider all stationary processes whose transition law is given by a time-homogeneous transition matrix P. The initial distribution must therefore be the stationary distribution π, which corresponds to P. The transition law of such a process is therefore given by a pair (π, P), where $\pi P = \pi$. It is clear that $\text{distr}(X_k) = \mathcal{E}$, where $\mathcal{E} = \{\pi \colon \exists P \in \mathcal{P}, \pi P = \pi\}$ is the set of all stationary distributions corresponding to the transition matrices in $\mathcal{P}$. From the general theory of reversible precise Markov chains, it follows that π is a stationary distribution corresponding to the reversed matrix P^* of P, and therefore the reverse process is one with the transition law given by (π, P^*). The reversed imprecise process is therefore the set of all processes of the given form.

One could also consider a set of processes with the set of transition laws (q, P), where q does not necessarily have to be a stationary distribution with respect to P. However, the reversals of such processes are generally not time-homogeneous. Such processes are therefore analysed in the context of the more general processes induced by the strong independence interpretation.

Under the strong independence interpretation, the imprecise Markov chain $(X_{1:n})$ is represented as a set of processes whose transition laws have the form $\Gamma = (q^1, P_{1:n-1})$, where the initial distribution q^1 belongs to $\mathcal{C}$ and the transition matrices P_k each belong to the set $\mathcal{P}$. The set of reversed processes is simply the set consisting of processes with transition laws $\Gamma^* = (q^n, P^*_{n-1:1})$.

In the context of epistemic irrelevance, the set of processes that follow an imprecise transition law generally does not exhibit the Markov property. This means that the state distributions at later times are generally affected by the overall history of the process, but still remain within the bounds set by the imprecise transition law. Consequently, one could marginalise the resulting joint distributions to create a reverse process. In this reverse process, the state distributions are influenced by future states and not by past states. Nevertheless, the methodology to analytically link a reverse process to the corresponding forward process, similar to strong independence case, remains elusive. Therefore, this topic is identified as an area for further investigation.

5.2 Properties of Reversed Imprecise Markov Chains

Consider an imprecise Markov chain where the transition law is given by the set of initial distributions $\mathcal{C}$ and the transition matrices $\mathcal{P}$. Under the strong independence interpretation, the set of compatible processes contains all processes of the form $\Gamma = (q^1, P_{1:n-1})$, where q^1 is an arbitrary element of $\mathcal{C}$ and P_k can be chosen arbitrarily from $\mathcal{P}$. However, more restrictions apply when considering the reversed process and the corresponding processes with transition laws $\Gamma^* = (q^n, P^*_{n-1:1})$. While the initial distributions q^n can still be chosen arbitrarily from the set $\mathcal{C}^n = \mathcal{C}\mathcal{P}^{n-1}$, the set of transition matrices P^*_k for k between 1 and $n-1$ for a given distribution q^{k+1} is restricted to those of the form P^*_q, where $qP = q^{k+1}$. In principle, this violates the Markov property, because the set of possible transition laws depends on the distribution over the states at a given point in time.

5.3 Reversibility

The concept of reversibility can be extended to the imprecise case in a natural way. It usually requires stationarity, which means that all marginal distribution sets corresponding to certain X_k are the same for all k. This requirement is fulfilled if the initial set $\mathcal{C}$ is a stationary set, i.e. $\mathcal{C}\mathcal{P} = \mathcal{C}$ is fulfilled. An imprecise Markov chain $(X_{1:n})$ is therefore *stationary* if all X_k have the same marginal distributions: $\mathrm{distr}(X_k) = \mathrm{distr}(X_l)$ for each $1 \leq k, l \leq n$. While the detailed balance condition is not clear to formulate in the imprecise case, we require the same distribution condition instead. Therefore, we call a stationary imprecise Markov chain *reversible* whenever the joint distribution of the reversed process is the same as that of the original chain. The following proposition is immediate.

Proposition 4. *A stationary imprecise Markov chain* $(X_{1:n})$ *is reversible if and only if* $\mathrm{distr}(X_{1:n}) = \mathrm{distr}(X_{1:n})^*$.

The above proposition characterizes reversible imprecise Markov chains. However, due to the nature of the reversed processes, which generally do not satisfy the Markov property, it is unclear how such processes are generally generated.

However, some special cases simplify the analysis considerably. The first special case is the case of an imprecise Markov chain, which is understood as a precise chain with unknown parameters. As analysed above, in this interpretation, the chain induces the set of all processes whose transition law is of the form (π, P), where π is a stationary distribution with respect to P. The reversed law of such a process is then (π, P^*). According to Proposition 4, the chain is always reversible if every reversed transition matrix P^* is in $\mathcal{P}$ when P is. If we denote $\mathcal{P}^* = \{P^* \colon P \in \mathcal{P}\}$, then we can say that a chain is reversible according to this interpretation whenever its transition set satisfies $\mathcal{P}^* = \mathcal{P}$.

5.4 Imprecise Markov Chains with Precise Marginals

Further, we characterize a family of reversible Markov chains under the interpretation of strong independence. Consider an imprecise Markov chain model where the stationary distribution is precise, say π, while the transition law $\mathcal{P}$ is imprecise (see e.g. [10]). Such a situation would occur in cases where the available information is sufficient to precisely estimate the marginal distribution, but not the joint distribution as well. It is clear that under such assumptions the stationary set actually consists of a single element π and $\pi P = \pi$ for each $P \in \mathcal{P}$. We assume the interpretation of strong independence here.

From the above results, it clearly follows that an imprecise Markov chain with the initial distribution π and the transition law $\mathcal{P}$ is stationary under the above assumptions. Consider a compatible process with the transition rule $\Gamma = (\pi, P_{1:n-1})$. Then its reverse is $\Gamma^* = (\pi, P^*_{n-1:1})$, where $P^*_k = (P_k)^*_\pi$ are the standard reverses. The following proposition clearly holds.

Proposition 5. *Let $(X_{1:n})$ be an imprecise Markov chain with precise stationary distribution π and imprecise transition set $\mathcal{P}$. Under the assumption of the strong independence interpretation, the chain is reversible if and only if $\mathcal{P}^* = \mathcal{P}$.*

6 Discussion

We have shown that it is possible to connect a reversed process to an imprecise Markov chain by reversing the compatible non-homogeneous Markov processes. Despite the symmetry between forward and reversed processes, it turns out that the reversed processes generally lose important properties, such as the Markov property, unless some rather restrictive additional assumptions are made. The result obtained raises the question of the meaning and properties of the reversed processes and how the absence of the Markov property can be related to natural phenomena. Another related area of research is imprecise processes that are reversible. In particular, their characterization and relation to real-world processes.

References

1. Aldous, D., Fill, J.A.: Reversible Markov chains and random walks on graphs, 2002. In: Unfinished Monograph, Recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html
2. Brooks, S., Gelman, A., Jones, G., Meng, X.-L.: Handbook of Markov Chain Monte Carlo. CRC Press (2011)
3. De Cooman, G., Hermans, F., Quaeghebeur, E.: Imprecise Markov chains and their limit behavior. Probab. Eng. Inf. Sci. **23**(4), 597–635 (2009). https://doi.org/10.1017/S0269964809990039
4. Gallager, R.G.: Discrete stochastic processes. Massachusetts Institute of Technology, OpenCourseWare (2011)
5. Hartfiel, D.J.: Markov Set-Chains. Springer, Berlin (1998)
6. Ikeda, S., Kubo, I., Yamashita, M.: The hitting and cover times of random walks on finite graphs using local degree information. Theor. Comput. Sci. 410(1), 94–100 (2009). https://doi.org/10.1016/j.tcs.2008.10.020
7. Kelly, F.P.: Reversibility and Stochastic Networks. Cambridge University Press (2011)
8. Norris, J.R.: Markov Chains. Cambridge University Press, Cambridge (1997)
9. Škulj, D.: Discrete time Markov chains with interval probabilities. Int. J. Approx. Reason. **50**(8), 1314–1329 (2009). https://doi.org/10.1016/j.ijar.2009.06.007
10. Škulj, D.: Random walks on graphs with interval weights and precise marginals. Int. J. Approx. Reason. **73**, 76–86 (2016). https://doi.org/10.1016/j.ijar.2016.02.008
11. Stroock, D.W.: An Introduction to Markov Processes. Graduate Texts in Mathematics, Vol. 230. Springer, Berlin (2013)
12. Trendelkamp-Schroer, B., Wu, H., Paul, F., Noé, F.: Estimation and uncertainty of reversible Markov models. J. Chem. Phys. **143**, 174101 (2015)

Minimum Intrinsic Dimension Scaling for Entropic Optimal Transport

Austin J. Stromme(✉)

ENSAE/CREST, 5 Av. Le Chatelier, 91120 Palaiseau, France
austin.stromme@ensae.fr

Abstract. Motivated by the manifold hypothesis, which states that data with a high extrinsic dimension may yet have a low intrinsic dimension, we develop refined statistical bounds for entropic optimal transport that are sensitive to the intrinsic dimension of the data. Our bounds involve a novel notion of intrinsic dimension, measured at only a single distance scale depending on the regularization parameter, and show that it is only the minimum of these single-scale intrinsic dimensions which governs the rate of convergence. We call this the Minimum Intrinsic Dimension scaling (MID scaling) phenomenon, and establish MID scaling with no assumptions on the data distributions so long as the cost is bounded and Lipschitz, and for various entropic optimal transport quantities beyond just values, with stronger analogs when one distribution is supported on a manifold. Our results significantly advance the theoretical state of the art by showing that MID scaling is a generic phenomenon, and provide the first rigorous interpretation of the statistical effect of entropic regularization as a distance scale. In this extended abstract, we state our main results and discuss MID scaling, while deferring proofs to the main article.

Keywords: Entropic optimal transport · Intrinsic dimension · Statistical rates

1 Introduction

Optimal transport (OT) is a powerful paradigm for comparing probability distributions based on a minimum-energy criterion, and has recently been employed throughout applied science, including statistics, computer science, biology and beyond [16,17,21,22,27]. In practice, the OT problem is most often regularized with an entropic penalty, leading to the entropic OT problem. For probability measures μ, ν on $\mathbb{R}^d$ and a cost function $c\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the entropic OT problem, with respect to the cost c and with regularization $\varepsilon > 0$, is defined as

$$S_\varepsilon(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int c(x,y)\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi \,\|\, \mu\otimes\nu), \tag{1}$$

where $\Pi(\mu,\nu)$ denotes the set of probability distributions π on $\mathbb{R}^d \times \mathbb{R}^d$ which couple μ to ν, namely such that their marginal on the first d coordinates is

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 491–499, 2024.
https://doi.org/10.1007/978-3-031-65993-5_60

μ and their marginal on the last d coordinates is ν, and $\mu \otimes \nu$ denotes the joint law of (x, y) when $x \sim \mu$ and $y \sim \nu$ are independent. This entropically regularized problem is typically preferred because it is amenable to Sinkhorn's algorithm [3,24,25], which offers quadratic scaling in the support size, simplicity, and parallelizability.

1.1 Statistical Aspects of (Entropic) OT

Unfortunately, the un-regularized OT problem $S_0(\mu, \nu)$ is known to suffer from a statistical curse of dimensionality. To describe this barrier, consider the practical situation in which one does not have access to the entire distributions μ, ν, and instead only has access to iid samples of size n from each distribution, which we write as $\mathcal{X} := (x_1, \ldots, x_n) \sim \mu^{\otimes n}$ and $\mathcal{Y} := (y_1, \ldots, y_n) \sim \nu^{\otimes n}$. Let $\hat{\mu}$ and $\hat{\nu}$ denote the empirical measures supported on $\mathcal{X}$ and $\mathcal{Y}$, respectively, namely $\hat{\mu} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, and $\hat{\nu} := \frac{1}{n}\sum_{j=1}^n \delta_{y_j}$. Given the empirical measures $\hat{\mu}, \hat{\nu}$, the most natural way to estimate the population quantity $S_0(\mu, \nu)$ is with the plug-in estimator $S_0(\hat{\mu}, \hat{\nu})$. The discrepancy between the empirical OT value and its population counterpart is an old and well-studied area, and it has long been understood that rates like $n^{-1/d}$ are typical [6], and it has recently been shown that such rates are essentially un-improvable [12,15].

Statistical entropic OT. Motivated by its computational benefits and corresponding prevalence in practice, as well as the curse of dimensionality for its un-regularized counterpart, a line of recent work has endeavored to understand the statistical consequences of entropic regularization [4,9,11,14,20]. This work has shown that the entropic OT problem (1) offers significantly improved statistical performance, essentially transferring the curse of dimensionality from the sample size n to the regularization parameter ε; the following result describes the current state of the art when the measures are bounded and ε is small.

Theorem 1 ([9,14]). *Suppose μ, ν are probability measures on $\mathbb{R}^d$ with support contained in the unit ball $B(0, 1)$. Then for a constant C_d depending only on the dimension d,*

$$\mathbb{E}[|S_\varepsilon(\hat{\mu}, \hat{\nu}) - S_\varepsilon(\mu, \nu)|] \leqslant C_d \cdot \frac{1 + \varepsilon^{-\lfloor d/2 \rfloor}}{\sqrt{n}},$$

where the expectation is taken over the iid samples $\mathcal{X} \sim \mu^{\otimes n}, \mathcal{Y} \sim \nu^{\otimes n}$.

By giving an explicit trade-off between regularization and statistical error, this result offers a strong and flexible description of the performance of entropic OT in practice. In this work, we build on Theorem 1 and develop a refined theory of the statistical behavior of entropic OT.

1.2 Effective Statistical Dimension of Entropic OT

To describe our refinements, observe that the estimate in Theorem 1 is *extrinsic* in the sense that the dimension d is appearing because the ambient space is $\mathbb{R}^d$, rather than anything to do with the *intrinsic* dimensions of μ and ν.

This extrinsic dimension dependence is fundamental to the proof, and can be very pessimistic. For a toy example, suppose μ and ν are supported on the same k-dimensional hyperplane in $\mathbb{R}^d$: then Theorem 1 itself implies that the statistical rate has the potentially much milder dependence $\varepsilon^{-\lfloor k/2 \rfloor}$ on the regularization parameter ε. In such a case, we can say that μ, ν have extrinsic dimension d yet have intrinsic dimension (at most) k. Well beyond toy settings, the widely believed manifold hypothesis states that natural data with a high extrinsic dimension is typically near to or on a low-dimensional manifold embedded in Euclidean space, and so has low intrinsic dimension [1,7]. Given the ubiquity of entropic OT in practice, it is therefore of major interest to determine the effective statistical dimension of entropic OT, and identify how it relates to the extrinsic dimension d as well as the intrinsic dimensions of μ and ν. In this work, we thus study the following question:

Suppose μ and ν are supported on $\mathbb{R}^d$, yet have (informally speaking) intrinsic dimensions d_μ and d_ν. How does the statistical rate of convergence of entropic OT depend on d, d_μ, and d_ν?

Our results will measure intrinsic dimension through the covering numbers of the supports of μ and ν. Recall that the covering number of a set $A \subset \mathbb{R}^d$ at scale $\delta > 0$ is defined as the minimum number of closed δ balls need to cover A, and written $\mathcal{N}(A, \|\cdot\|, \delta)$.

It will be convenient to define for all $\delta > 0$

$$\mathcal{N}(\mu, \delta) := \mathcal{N}(\mathrm{supp}(\mu), \|\cdot\|, \delta), \qquad \mathcal{N}(\nu, \delta) := \mathcal{N}(\mathrm{supp}(\nu), \|\cdot\|, \delta). \tag{2}$$

Contributions

Using this notation, we can give our answer to the above question:

MID scaling: *the statistical rate of convergence of entropic OT can be upper bounded by quantities whose dimension-dependence is purely contained within the minimum covering number at scale ε, namely $\mathcal{N}(\mu, \varepsilon) \wedge \mathcal{N}(\nu, \varepsilon)$.*

We call this minimum intrinsic dimension scaling (MID scaling). We emphasize that MID scaling encapsulates two related yet distinct phenomena:

1. **Minimum**: only the *minimum* of the intrinsic dimensions governs the convergence
2. **Intrinsic Dimension**: the dimension-dependence is intrinsic to μ and ν, and in fact is intrinsic *at a single distance scale depending on the regularization parameter.*

The above statement of MID scaling is in the setting of Theorem 1, but our main results establish MID scaling more generally for bounded, Lipschitz costs, and for entropic OT quantities beyond values, including maps and densities.

1.3 Preliminaries

Throughout, we work under the following assumption on the cost function c.

Assumption 1 (Bounded and uniformly Lipschitz cost). We assume the cost c is Borel measurable and $c \in L^\infty(\mu \otimes \nu)$. By re-scaling the problem we assume without loss of generality that $\|c\|_{L^\infty(\mu\otimes\nu)} \leqslant 1$. Also, we assume there exists $L > 0$ such that c is L-Lipschitz in each variable, uniformly over $\mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$.

There are unique optimal solutions to the problems $S_\varepsilon(\mu, \nu)$ and $S_\varepsilon(\hat{\mu}, \nu)$, which we write as $\pi_\varepsilon \in \Pi(\mu, \nu)$ and $\hat{\pi}_\varepsilon \in \Pi(\hat{\mu}, \hat{\nu})$, respectively.

*Strong duality.*Under Assumption 1 on the cost c, a form of strong duality holds for both the population and empirical entropic OT problems [5]. For the population entropic OT problem, strong duality is

$$S_\varepsilon(\mu, \nu) = \sup_{(f,g)\in L^\infty(\mu)\times L^\infty(\nu)} \Phi_\varepsilon(f, g) := \mu(f) + \nu(g) - \varepsilon(\mu \otimes \nu)(e^{-\varepsilon^{-1}(c-f-g)} - 1).$$

The function $\Phi_\varepsilon \colon L^\infty(\mu) \times L^\infty(\nu) \to \mathbb{R}$ is the population entropic OT dual function. Similarly for the empirical entropic OT problem, strong duality is stated

$$S_\varepsilon(\hat{\mu}, \hat{\nu}) = \sup_{(f,g)\in L^\infty(\hat{\mu})\times L^\infty(\hat{\nu})} \hat{\Phi}_\varepsilon(f, g) := \hat{\mu}(f) + \hat{\nu}(g) - \varepsilon(\hat{\mu} \otimes \hat{\nu})(e^{-\varepsilon^{-1}(c-f-g)} - 1).$$

And this $\hat{\Phi}_\varepsilon \colon L^\infty(\hat{\mu}) \times L^\infty(\hat{\nu}) \to \mathbb{R}$ is the empirical entropic OT dual function.

Population and empirical dual potentials.

The population and empirical dual problems have solutions $(f_\varepsilon, g_\varepsilon) \in L^\infty(\mu) \times L^\infty(\nu)$ and $(\hat{f}_\varepsilon, \hat{g}_\varepsilon) \in L^\infty(\hat{\mu}) \times L^\infty(\hat{\nu})$, respectively, which are uniquely specified by assuming $\nu(g_\varepsilon) = 0$ and $\hat{\nu}(\hat{g}_\varepsilon) = 0$.

*Population and empirical densities.*For notational convenience, we also define the population density, for μ-almost every x and ν-almost every y, as

$$p_\varepsilon(x, y) := \frac{\mathrm{d}\pi_\varepsilon}{\mathrm{d}(\mu \otimes \nu)}(x, y) = e^{-\varepsilon^{-1}(c(x,y) - f_\varepsilon(x) - g_\varepsilon(y))}. \tag{3}$$

And we similarly define, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the empirical density

$$\hat{p}_\varepsilon(x, y) := \frac{\mathrm{d}\hat{\pi}_\varepsilon}{\mathrm{d}(\hat{\mu} \otimes \hat{\nu})}(x, y) = e^{-\varepsilon^{-1}(c(x,y) - \hat{f}_\varepsilon(x) - \hat{g}_\varepsilon(y))} \tag{4}$$

2 Main Results

2.1 MID Scaling with Value Convergence

Recall for the covering numbers of the support of μ and ν from (2). Our main result on entropic OT value estimation follows.

Theorem 2 (MID scaling for values). *For numerical constants independent of all problem parameters,*

$$\mathbb{E}[|S_\varepsilon(\hat{\mu},\hat{\nu}) - S_\varepsilon(\mu,\nu)|] \lesssim (1+\varepsilon)\sqrt{\frac{\mathcal{N}(\mu,\frac{\varepsilon}{L}) \wedge \mathcal{N}(\nu,\frac{\varepsilon}{L})}{n}}.$$

The only dimensional quantity in this estimate is contained in the minimum covering numbers at scale ε, demonstrating the MID scaling phenomenon. We emphasize that this result only requires that the cost c is bounded and Lipschitz (Assumption 1), in contrast to the smoothness assumptions in most of the previous literature. Theorem 2 improves upon Theorem 1, up to a factor of $\sqrt{\varepsilon}$ in odd dimension, under weaker assumptions on the cost (smooth vs. Lipschitz).

2.2 MID Scaling for Maps and Densities

Theorem 2 gives a strong instance of the MID scaling phenomenon for entropic optimal transport, yet in many applications of entropic OT, such as trajectory reconstruction [22] and domain adaptation [8,23], it is of greater interest to estimate the solution to the entropic OT problem than the value of the entropic OT objective alone. Motivated by this fact, we also study the performance of empirical plug-in estimators for the problems of entropic OT map estimation as well as density estimation, and develop results analogous to Theorem 2, showing that natural plug-in estimators also sport MID scaling.

We define the entropic OT map $T_\varepsilon(x) := \mathbb{E}_{\pi_\varepsilon}[y \,|\, x]$. The map T_ε is an entropic analog of the OT map, and has been the subject of much recent work as it offers greater computational and statistical efficiency than its un-regularized counterpart [13,18–20]. We study the empirical analog of $T_\varepsilon(x)$, namely $\hat{T}_\varepsilon(x) := \mathbb{E}_{\hat{\pi}_\varepsilon}[y \,|\, x]$, and show that it attains a $1/\sqrt{n}$ rate, with MID scaling.

Theorem 3. *Suppose the diameter of* $\mathrm{supp}(\nu)$ *is at most* R*. Then for numerical constants independent of all problem parameters,*

$$\mathbb{E}[\|\hat{T}_\varepsilon - T_\varepsilon\|^2_{L^2(\hat{\mu})}] \lesssim R^2\big(1+\frac{1}{\varepsilon}\big)\sqrt{\frac{\mathcal{N}(\mu,\frac{\varepsilon}{L}) \wedge \mathcal{N}(\nu,\frac{\varepsilon}{L})}{n}}.$$

We note that this result is measured with respect to the empirical norm $\|\cdot\|_{L^2(\hat{\mu})}$, rather than the population norm $\|\cdot\|_{L^2(\mu)}$. In the next section we assume that one distribution is supported on an embedded manifold and show that this result can be strengthened to population norm convergence with fast $1/n$ rates. Previous work on the convergence of $\hat{T}_\varepsilon$ to T_ε for generic distributions incurred an exponential dependence on $1/\varepsilon$ [13,20]. And an alternative estimator which achieves sub-exponential dependendence on $1/\varepsilon$ was considered in [26], but at the cost of worse rates in n. Theorem 2 avoids exponential factors and achieves MID scaling with a $1/\sqrt{n}$ rate.

MID scaling for entropic OT density estimation. We can also consider estimating the full entropic OT density p_ε, defined in (3). A natural way to estimate p_ε is with its plug-in counterpart, $\hat{p}_\varepsilon$ from (4). We obtain the following result.

Theorem 4. *For numerical constants independent of all problem parameters,*

$$\mathbb{E}[\|\hat{p}_\varepsilon - p_\varepsilon\|_{L^1(\hat{\mu}\otimes\hat{\nu})}] \lesssim \big(1 + \frac{1}{\sqrt{\varepsilon}}\big)\Big(\frac{\mathcal{N}(\mu, \frac{\varepsilon}{L}) \wedge \mathcal{N}(\nu, \frac{\varepsilon}{L})}{n}\Big)^{1/4}.$$

This result shows that MID scaling even applies when estimating the full density p_ε, albeit in the empirical $L^1(\hat{\mu} \otimes \hat{\nu})$ norm. Previously known bounds for the convergence of $\hat{p}_\varepsilon$ to p_ε incurred an exponential dependence on $1/\varepsilon$ [20], and Theorem 4 removes this dependence while only incurring MID scaling.

2.3 Fast Rates with MID Scaling on Embedded Manifolds

In this section, we describe how we can strengthen the results from the previous section when the low-dimensional measure is supported on a manifold.
Assumptions. We work under the following assumptions, and for simplicity assume that ν is the low-dimensional measure.

Assumption 2 (ν is supported on an embedded manifold). Assume ν is supported on a compact, smooth, connected, Riemannian manifold (N, h) of dimension $d_\nu \geqslant 3$ without boundary, where N is endowed with the submanifold geometry from its inclusion in $\mathbb{R}^d$.

Assumption 3 (ν is Lipschitz and non-vanishing). Assume that ν has a Lipschitz, non-vanishing density with respect to the Riemannian volume on N.

Main result on dual potential convergence. The purpose of this section is to show that under these assumptions on ν and N, we can derive significantly stronger instances of MID scaling than in the previous section. Our main result concerns the convergence of the empirical dual potentials $\hat{f}_\varepsilon, \hat{g}_\varepsilon$ to their population counterparts $f_\varepsilon, g_\varepsilon$. Such convergence is essentially stronger than that of entropic OT values, maps, and densities, since each of those quantities is defined in terms of the dual potentials.

Theorem 5. *Make Assumptions 1, 2, and 3. If ε/L is sufficiently small, then, suppressing constants depending on N and ν but not the extrinsic dimension d, we have*

$$\mathbb{E}\big[\|\hat{f}_\varepsilon - f_\varepsilon\|^2_{L^2(\mu)} + \|\hat{g}_\varepsilon - g_\varepsilon\|^2_{L^2(\nu)}\big] \lesssim \big(\varepsilon^2 + \frac{1}{\varepsilon^2}\big) \cdot \Big(\frac{L}{\varepsilon}\Big)^{13d_\nu+8} \cdot \frac{1}{n}.$$

To the best of our knowledge, all previously known bounds on entropic OT dual potentials incurred an exponential dependence on ε [4,11,20]. The significance of Theorem 5 is that it avoids exponential dependence on ε while only incurring MID scaling dependence. See the full paper for applications of this result to obtain stronger analogs of the previous results.

2.4 Conclusion and Discussion of MID Scaling

MID scaling is similar to, but distinct from, the intrinsic dimension-dependence in the un-regularized OT literature [2,6,10,28]. While in both settings the minimum of the intrinsic dimensions governs the rate of convergence, the notion of intrinsic dimension is different. In the un-regularized setting, intrinsic dimension is characterized through uniform covering number control at small scales, whereas in MID scaling the covering numbers play a role, but only at a single distance scale. The convergence of un-regularized OT is, in fact, adaptive to multi-scale behavior, but the relevant scales are determined by the sample size n, meaning that milder covering numbers at some scales only translate to improved rates while n is not too large; in fact, such rates are known to be tight [28]. In contrast, MID scaling shows that entropic OT benefits from better covering number control at some scales *for all sample sizes* n. Essentially, entropic regularization decouples the sample size and the distance scale of the problem, allowing for a flexible trade-off between the intrinsic dimension of the data distribution and the amount of regularization.

MID scaling helps clarify the statistical role of entropic regularization, showing that beyond its well-known computational virtues, entropy also provides statistical regularization by specifying a distance scale, allowing the user to balance the intrinsic curse of dimensionality of the data with the statistical difficulty of the problem. Because of the manifold hypothesis, which states that natural data is typically supported on or near a low-dimensional embedded manifold, we expect data to have significantly smaller intrinsic dimension than extrinsic dimension. And so a means to flexibly adapt optimal transport to such intrinsic low-dimensional structure, particularly approximate low-dimensional structure, is of major interest. MID scaling demonstrates that entropic regularization provides this benefit for optimal transport.

References

1. Bengio, Y., Goodfellow, I., Courville, A.: Deep Learning, vol. 1. MIT Press, Cambridge (2017)
2. Boissard, E., Le Gouic, T.: On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. Ann. Inst. H. Poincaré Probab. Statist. **50**, 539–563 (2014)
3. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems (NIPS 2013), Vol. 26 (2013)
4. del Barrio, E., González-Sanz, A.G., Loubes, J.-M., Niles-Weed, J.: An improved central limit theorem and fast convergence rates for entropic transportation costs. SIAM J. Math. Data Sci. **5**(3), 639–669 (2022)
5. Di Marino, S., Gerolin, A.: An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn Algorithm. J. Sci. Comput. **85**, 27 (2020)
6. Dudley, R.M.: The speed of mean Glivenko-Cantelli convergence. Ann. Math. Stat. **40**(1), 40–50 (1969)

7. Fefferman, C., Mitter, S., Narayanan, H.: Testing the manifold hypothesis. J. Am. Math. Soc. **29**(4), 983–1049 (2016)
8. Flamary, R., Courty, N., Tuia, D., Rakotomamonjy, A.: Optimal transport for domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(9), 1853–1865 (2017)
9. Genevay, A., Chizat, L., Bach, F., Cuturi, M., Peyré, G.: Sample complexity of Sinkhorn divergences. In: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, Vol. 89, pp. 1574–1583 (2019)
10. Hundrieser, S., Staudt, T., Munk, A.: Empirical Optimal Transport Between Different Measures Adapts To Lower Complexity. arXiv:2202.10434 (2022)
11. Luise, G., Salzo, S., Pontil, M., Ciliberto, C.: Sinkhorn Barycenters with free support via Frank-Wolfe Algorithm. In: Advances in Neural Information Processing Systems (NeurIPS 2019), Vol. 32 (2019)
12. Manole, T., Balakrishnan, S., Niles-Weed, J., Wasserman, L.: Plugin Estimation of Smooth Optimal Transport Maps. arXiv:2107.12364v1 (2021)
13. Masud, S.B., Werenski, M., Murphy, J.M., Aeron, S.: Multivariate Rank Via Entropic Optimal Transport: Sample Efficiency and Generative Modeling. arXiv:2111.00043v1 (2021)
14. Mena, G., Niles-Weed, J.: Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. In: Advances in Neural Information Processing Systems (NeurIPS 2019), Vol. 32 (2019)
15. Niles-Weed, J., Rigollet, P.: Estimation of Wasserstein distances in the spiked transport model. Bernoulli **28**(4), 2663–2688 (2022)
16. Panaretos, V.M., Zemel, Y.: Statistical aspects of Wasserstein distances. Annu. Rev. Stat. Appl. **6**, 405–431 (2019)
17. Peyré, G., Cuturi, M.: Computational optimal transport. Found. Trends Mach. Learn. **11**(5–6), 355–607 (2019)
18. Pooladian, A.-A., Divol, V., Niles-Weed, J.: Minimax estimation of discontinuous optimal transport maps: the semi-discrete case. In: Fortieth International Conference on Machine Learning (ICML 2023) International, pp. 28128–28150. PMLR (2023)
19. Pooladian, A.-A., Niles-Weed, J.: Entropic Estimation of Optimal Transport Maps. arXiv:2109.12004v1 (2021)
20. Rigollet, P., Stromme, A.J.: On the Sample Complexity of Entropic Optimal Transport. arXiv:2206.13472 (2022)
21. Santambrogio, F.: Optimal transport for applied mathematicians. In: Progress in Nonlinear Differential Equations and their Applications, Vol. 87. Birkhäuser/Springer, Cham (2015)
22. Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., et al.: Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell **176**(4), 928–943 (2019)
23. Seguy, V.,Damodaran, B.B., Flamary, R., Courty, N., Rolet, A., Blondel, M.: Large-scale optimal transport and mapping estimation. In: ICLR 2018—International Conference on Learning Representations (2018)
24. Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann. Math. Stat. **35**(2), 876–879 (1964)
25. Sinkhorn, R., Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices. Pacific J. Math. **23**(3), 343–348 (1967)

26. Stromme, A.J.: Sampling from a Schrödinger bridge. In: Proceedings of The 26th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, Vol. 206, pp. 4058–4067 (2023)
27. Villani, C.: Optimal Transport: Old and New. Springer Science & Business Media (2008)
28. Weed, J., Bach, F.: Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. Bernoulli **25**(4A), 2620–2648 (2019)

Creation of Probabilistic Models by Experts: A Case Study

Radek Švadlenka[1(✉)] and Radim Jiroušek[2]

[1] Faculty of Management, Prague University of Economics and Business, Jindichv Hradec, Czech Republic
Radek.Svadlenka@alef.com

[2] Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czech Republic
radim@utia.cas.cz

Abstract. Several approaches for the data-based construction of probabilistic models have been described in the literature. The present paper presents an alternative approach based on experts' knowledge. Naturally, the knowledge must be somehow transformed into probabilistic notions, which are, in our case, two-dimensional contingency tables. The basic idea is that even non-mathematicians can easily understand simple probabilistic elements (like two-dimensional distributions), from which one can assemble non-trivial complex models. For this, we use decomposable compositional models. The resulting models may be equivalent to a classical Bayesian network or a causal model of Pearl's type if the experts can determine which variables are causes and which are their consequences.

Keywords: Probability distribution · Graphical model · Composition · Mutual information · Information content · Maximum entropy principle

1 Organization's Cybersecurity

There are many security standards, methodologies, and legislation at the local and European levels, and new standards are constantly being prepared. To estimate the impact of newly introduced measures beforehand, one needs a model describing the interdependencies among the main features (characteristics) influencing the cybersecurity of big organizations that may be a target of attackers and foreign intelligent services. To employ data-based model learning approaches ready-made in several software packages, one needs data that are not available in a sufficient amount because such data are sensitive, and each organization keeps them secret. The paper describes an experiment to construct a model based on the knowledge of a group of experts. There are too few of them to supply representative data, but quite a lot of them, so their joint expertise may guarantee the validity of the resulting model.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 500–508, 2024.
https://doi.org/10.1007/978-3-031-65993-5_61

2 Expert Knowledge Formalization

In the described case, the knowledge acquisition process started with structured discussions with top security and IT managers of 52 large organizations from the public and commercial sectors. The achieved results exploited for the design of the model structure are twofold:

- The experts concurred that the model should describe organizations with six characteristics (in a model, we will call them random variables) listed in Table 1. L is dichotomic, and all other variables show to what extent the organizations meet the predefined conditions using a scale $1-5$ (or, $1-4$).
- Each top executive submitted the self-assessment[1] of their organization using the features from Table 1.

Table 1. Features (variables) characterizing organizations

Denotation	Variable	Number of values
L	Legislation (regulated/unregulated by law)	2
S	Security knowledge status of decision makers	5
T	Total security score of organization	5
U	User experience	4
V	Volume of resources invested in cybersecurity	5
W	Willingness of decision makers to educate themselves	5

Thus, each security manager described their organization by a vector of values of six variables L, S, T, U, V, W, which formed the basic knowledge for constructing the model describing the mutual relations among the variables.

2.1 Information-Theoretic Measures

Initially, we exploited the formalized self-assessments to evaluate the strength of mutual dependence of pairs of variables. For each couple of variables, we set up a contingency table as that in Table 2 and computed the corresponding mutual information according to the following formula[2]

$$MI(S,W) = \frac{1}{n}\sum_{i,j} n_{ij} \cdot \log_2\left(n \cdot \frac{n_{ij}}{n_{i.} \cdot n_{.j}}\right). \tag{1}$$

[1] Naturally, the authors committed not to publish any sensitive data, especially those making identification of the engaged organizations and their security managers' self-assessments possible.

[2] n_{ij} denotes the number appearing on the i-th row and j-th column in the contingency table. $n_{i.}$ and $n_{.j}$ are the corresponding row and column sums, respectively. n is a total sum in the contingency table, i.e., for Table 2, $n = 52$.

Table 2. Contingency table for variables S and W

	Variable S					$n_{i.}$
Variable W	8	2	0	0	0	10
	1	5	2	2	0	10
	0	3	2	5	1	11
	0	3	2	2	2	9
	0	0	0	2	10	12
$n_{.j}$	9	13	6	11	13	$n = 52$

Recall some elementary properties of the mutual information [3]. It measures the dependence of the considered pair of variables; the stronger the dependence, the higher the corresponding mutual information. It is non-negative, and only for independent variables it equals 0. It is always lower than Shannon's entropy of both the considered variables. Therefore, we often prefer considering its normalized version, so-called *information measure of dependence* defined

$$IMD(S,W) = \frac{MI(S,W)}{\min(H(n_{i.}), H(n_{.j}))}, \tag{2}$$

where

$$H(n_{i.}) = \frac{-1}{n} \sum_i n_{i.} \cdot \log_2 \left(\frac{n_{i.}}{n}\right), \qquad H(n_{.j}) = \frac{-1}{n} \sum_j n_{.j} \cdot \log_2 \left(\frac{n_{.j}}{n}\right).$$

The values of mutual information and the information measure of dependence for the considered variables are listed (in a descendant order according to IMD) in Table 3. Because the considered simple model could not incorporate all the dependencies, this auxiliary table helped the experts to choose the most important relationships that should be incorporated into the resulting model. The requirement was that each variable was selected at least once, and the pairs of selected variables corresponded to the edges of a decomposable graph (see the emphasized values).

Table 3. Values of information measures for pairs of variables

	WS	ST	TU	WT	VU	VT	SU	WL	WU	WV	SV	SL	TL	UL	VL
MI	.924	.737	.450	.485	.416	.377	.301	.132	.266	.265	.241	.096	.090	.043	.040
IMD	**.407**	**.319**	**.227**	**.213**	**.209**	**.169**	.152	**.135**	.134	.119	.108	**.098**	.092	.044	.041

2.2 Compositional Probabilistic Models

A general probability distribution for our six variables is defined by $2 \cdot 5^4 \cdot 4 - 1 = 4\,999$ parameters (probabilities). Determining so many parameters is beyond the

scope of any experts. This is why we consider a class of distributions defined by fewer parameters.

A *compositional model* is a probability distribution that is *composed* from a sequence of its low-dimensional marginal distributions by the operator of composition [6]. In this paper, we consider only *graphical* compositional models, the marginals of which can be represented by cliques of a graph. Considering the selected pairs of variables (see emphasized values in Table 3) as edges of a graph in Fig. 1, the corresponding compositional models are six-dimensional probability distributions for which

$$\begin{aligned}\pi(L,S,T,U,V,W) &= \pi(S,T,W) \triangleright \pi(T,U,V) \triangleright \pi(L,S,W) \\ &= \pi(S,T,W) \cdot \pi(U,V|T) \cdot \pi(L|S,W).\end{aligned} \tag{3}$$

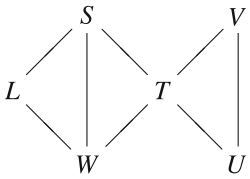


Fig. 1. Decomposable graph with three cliques: $\{L,S,W\}$, $\{S,T,W\}$, and $\{T,U,W\}$

In what follows, we use the following notation. (For more general definitions and advanced properties of this apparatus, see [6].) Consider a probability distribution π defined for a group of variables r. If $s \subsetneq r$, then $\pi^{\downarrow s}$ denote the marginal distribution of π restricted to variables s. Realize that $\pi^{\downarrow \emptyset} = 1$.

Let π and κ be distributions defined for variables r and t, respectively. Their composition is a distribution $(\pi \triangleright \kappa)$ for variables $r \cup t$

$$(\pi \triangleright \kappa) = \frac{\pi \cdot \kappa}{\kappa^{\downarrow r \cap t}}, \tag{4}$$

if the right-hand-side formula in Eq. (4) is defined. Notice that the composition of two marginals of a multidimensional probability distribution is always defined. It is also important to recall from [6] that this operator is generally not commutative nor associative. Therefore, to make the expressions like that in Eq. (3) unambiguous, the operators are always applied from left to right (if not explicitly changed by parentheses).

Recall from graph theory that a simple graph with cliques $c_1, c_2, \ldots, c_k$ is said to be *decomposable* (or *triangulated*) if the cliques can be enumerated so that they meet the so-called *running intersection property* (RIP):

$$\forall i = 3,4,\ldots,k \;\; \exists j \; (1 \le j < i) : c_i \cap (c_1 \cup \ldots \cup c_{i-1}) \subseteq c_j.$$

Probability distribution π for variables r is said to be *decomposable* if there exists a decomposable graph with cliques $c_1, \ldots, c_k$ such that $\pi = \pi^{\downarrow c_1} \triangleright \ldots \triangleright \pi^{\downarrow c_k}$, if the ordering $c_1, \ldots, c_k$ meets RIP.

Note that it was shown that for decomposable distribution π, $\pi = \pi^{\downarrow c_1} \triangleright \ldots \triangleright \pi^{\downarrow c_k}$ for all RIP orderings of the cliques $c_1, \ldots, c_k$ [6]. When designing the required model, we will also employ the following statement.

Proposition 1. *Let π be a decomposable distribution with the decomposable graph G, the cliques of which are $c_1, \ldots, c_k$. Consider any probability distribution κ of variables $\bigcup_{j=1}^{k} c_j$. If $\pi^{\downarrow j} = \kappa^{\downarrow j}$ for all $j = 1, \ldots, k$, then $H(\kappa) \leq H(\pi)$.*

2.3 Pairwise Dependence Corrections

As said above, based on values in Table 3, the experts selected eight pairs of variables, the dependence of which should mirrored in the resulting model. If we directly accepted the relative frequencies from the contingency table in Table 2 for the marginal probabilities of the model, it could lead to what the machine learning community calls *overlearning*. Such a model would not cover organizations for which, e.g., $W = 5$ and $S = 3$. Though none of the experts who contributed to this research assigned this combination of values to their organization, they admit that such a combination can occur for others. They want the model to admit this fact. They require the model to express a small quantity of ignorance.

Table 4. Modification for contingency table for variables S and W

		Variable S					
		1	2	3	4	5	$n_{i.}$
Variable W	1	8+ε	2+ε	ε	ε	0	10+4ε
	2	1+ε	5+ε	2+ε	2+ε	ε	10+5ε
	3	ε	3+ε	2+ε	5+ε	1+ε	11+5ε
	4	ε	3+ε	2+ε	2+ε	2+ε	9+5ε
	5	0	ε	ε	2+ε	10+ε	12+4ε
	$n_{.j}$	9+4ε	13+5ε	6+5ε	11+5ε	13+4ε	$n = 52 + 23\varepsilon$

To oblige this requirement, we slightly modified all eight contingency tables. Let us illustrate the procedure with Table 2. First, we asked the experts to determine whether there are zeroes in the contingency table that should be kept in the model, i.e., whether there are some logical relations. Regarding the contingency table describing the dependence between S and W, they concluded that, in their opinion, all zeroes should be eliminated but those expressing extreme combinations at the right-upper and left-lower corners. They did not expect that a manager declining any education ($W = 1$) could keep their knowledge status on the top level ($S = 5$), and, similarly, the knowledge of a manager eager to learn everything new would be as that of a beginner. Thus, the modification of this contingency table consisted in determining a positive ε that would yield the contingency table (see Table 4) meeting the experts' requirements best. To this end,

we do not have any theoretical support. Nevertheless, to help the experts decide on the value ε, we computed the values of mutual information and information measure of dependence for different ε; see Table 5. Finally, accepting the loss of 15.3 % of the mutual information between the considered variables, the experts chose $\varepsilon = 0.15$; a higher value of ε would waste too much information, and lower values would unnecessarily focus only on organizations like those taking part in the research.

Let us note that when realizing similar minor modifications of the remaining two-dimensional contingency tables, we kept all the tables pairwise consistent. It means that the one-dimensional marginals are the same no matter which of the two-dimensional tables they are computed from. For example, the marginal for variable T is the same for all the contingency tables for (S, T), or (T, U), (W, T), and (T, V). It is necessary to guarantee the convergence of the procedure described in the following section.

Table 5. Decrease of IMD with increasing ε in Table 4

ε	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	. 0.45	0.50
IMD	.407	.380	.361	.344	.330	.317	.305	.294	.283	.274	.265
Δ in %	0	6.4	11.2	15.3	19.9	22.1	25.1	27.8	30.3	32,6	34.7

2.4 Iterative Proportional Fitting Procedure

The procedure was published as early as in 1940 [5] (though its convergence was proved much later by Csiszár [4]), and, having already defined the composition operator, we can describe it easily.

Let $\pi_1, \pi_2, \ldots, \pi_k$ be probability distributions defined for groups of variables $c_1, c_2, \ldots, c_k$, respectively. Their ordering is arbitrary. As a first step (initialization), define the uniform probability distribution κ_0 for variables $\bigcup_{j=1}^{k} c_j$. The iterative proportional procedure consists of the repetitive computation

$$\kappa_i = \pi_{(i \bmod(k))} \rhd \kappa_{i-1}.$$

In other words, we take successively distributions $c_1, c_2, \ldots, c_k$, and compute their composition with the result of the preceding step. Each time, after getting to the end of the sequence, c_k, we start again with c_1 and continue until the sequence converges with the required precision. It directly follows from the results presented in [4] that the Shannon entropy of the limit distribution $H(\lim_{i\to\infty} \kappa_i)$ is greater or equal than the entropy of any distribution π (defined for variables $\bigcup_{j=1}^{k} c_j$) with all $\pi_1, \pi_2, \ldots, \pi_k$ for its marginals. This statement, in combination with Proposition presented in Sect. 2.2, yields that the six-dimensional decomposable model (with the graph from Fig. 1), the three-dimensional parts of which are computed from two-dimensional distributions using IPFP, is the maximal entropy extension of all the selected two-dimensional distributions.

3 Application

The preceding section described how we constructed the six-dimensional decomposable compositional model (see Eq. 3). We chose the edges of a decomposable graph based on mutual information values. Then, to prevent overlearning of the model, we slightly modified the corresponding two-dimensional contingency tables and eventually found the maximal entropy distributions $\pi(S,T,W)$, $\pi(T,U,V)$, $\pi(L,S,W)$ forming the model $\pi(L,S,T,U,V,W)$ according to Eq. (3). Using this model expressed, we can apply the efficient computational procedures described in [2] to compute any required marginal distribution or conditional probability.

To illustrate a possible application of the model, let us study how security regulation affects organizational behavior regarding managers' willingness to learn and the organization's overall level of security. Specifically, let us have a look at the behavior of managers with low to medium willingness to educate themselves in the field of cybersecurity (i.e., $W \leq 3$), who make decisions in organizations with a low to medium level of security (i.e., $S \leq 3$). The model shows substantial differences between the two types of organizations. Computing $\pi(S,W|L)$ we get

$$\pi(S \leq 3, W \leq 3|L = 1) = 0.27, \qquad \pi(S \leq 3, W \leq 3|L = 2) = 0.51.$$

This means that managers in lower security organizations regulated by law are more motivated to take cybersecurity training than those in unregulated organizations.

With this observation, a natural question arises: Can we expect that after a new law is approved and many more organizations will become regulated, the appetite for cybersecurity knowledge will also increase among top executives?

The answer to this question is complex, even if we assume the model is the best possible. The model describes the current situation, and the computation of conditionals given $L = 2$ means focussing on regulated organizations under the current state of law. It means that, as described by Pearl in [8], from this model, one cannot estimate the impact of introducing a new law, which moves some organizations from one category ($L = 1$) to another ($L = 2$). The asked question, what will happen after the new law comes into effect, could be answered not by conditioning ($\pi(\ldots|L = 2)$) but using the Pearl's *do operator* expressed $\pi(\ldots|do(L = 2))$, which may be different from ($\pi(\ldots|L = 2)$). However, it cannot be computed from the model we constructed but only from a *causal model*.

To explain the difference between conditional and the computation of the impact of an *intervention*, let us introduce a trivial causal example. Consider a contingency table describing the relationship between a fire alarm and smoke in a room. Computing the impact of intervention $\pi(alarm = on|do(smoke = yes)) = \pi(alarm = on|smoke = yes)$ because if we make smoke in a room (e.g., smoking a cigar), there is a high probability the alarm gets on. Contrary, $\pi(smoke = yes|alarm = on) \gg \pi(smoke = yes|do(alarm = on))$. The conditional probability $\pi(smoke = yes|alarm = on)$ is high because the alarm is usually on

when somebody smokes in the room. However, $\pi(smoke = yes|do(alarm = on))$

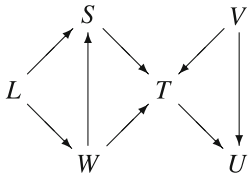


Fig. 2. Directed graph of a causal model

is the probability of smoke in a room when we make the alarm wail. The sound of a wailing alarm does not make smoke in a room.

Based on the constructed compositional model (Eq. 3), we studied the possibility of considering this model as a causal model of Pearl's type (i.e., without feedback). Realize that a causal model is represented identically as a Bayesian network. The corresponding graphs are directed and acyclic; however, the orientation of edges in a causal model must correspond to causality: the arrows must point from causes to consequences. The only difference is how we use it. The same also holds for compositional models, where the causal relations are encoded in the ordering in which low-dimensional distributions are composed into a multidimensional model. The causes must always precede their impacts; for more details see [7].

A weak point of the following consideration is that none of the experts was sure about any of the considered causal relations. Nevertheless, eventually, they agreed that assuming the relations are causal, they could be as expressed as in Fig. 2.

Thus, the following considerations are only speculative, assuming the graph in Fig. 2 describes the causal relations correctly. To encode the orientation of edges from Fig. 2 into the causal compositional model, we have to change the ordering of three-dimensional marginal distributions into the following one:

$$\pi_c(L, S, T, U, V, W) = \pi(L, S, W) \triangleright \pi(S, T, W) \triangleright \pi(T, U, V). \tag{5}$$

Two models, π defined by Eq. 3 and π_c defined by Eq. 5 in many aspects coincide; when computing conditionals, we get identical results. However, not to violate the causal relations when computing marginals, one has to be much more careful for π_c. Moreover, causal model π_c allows for computing the result of an intervention. Performing the required computations, we get

$$\pi_c(S \leq 3, W \leq 3|L = 2) = \pi_c(S \leq 3, W \leq 3|do(L = 2)) = 0.51$$

because variable L is at the beginning of the causal chain. This fact could raise hopes that the transposition of the NIS2 Directive into Czech legislation will positively impact the overall state of cybersecurity in the affected organizations.

Acknowledgments. The authors sincerely thank the 52 anonymous experts who devoted their time and shared their expertise.

References

1. Bína, V., Jiroušek, R.: On computations with causal compositional models. Kybernetika **51**, 525–539 (2015)
2. Bína, V., Jiroušek, R., Kratochvíl, V.: Foundations of compositional models: inference. Int. J. Gener. Syst. **50**, 409–433 (2021)
3. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley, New York (1990)
4. Csiszár, I.: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. **3**, 146–158 (1975)
5. Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. **11**, 427–444 (1940)
6. Jiroušek, R.: Foundations of compositional model theory. Int. J. Gen Syst **40**, 623–678 (2011)
7. Jiroušek, R., Shenoy, P.P.: Causal compositional models in valuation-based systems with examples in specific theories. Int. J. Approx. Reason. **72**, 95–112 (2016)
8. Pearl, J.: Causality. Cambridge University Press (2009)

Possibilistic Variables and Akian Semirings

Pedro Terán(✉)

Universidad de Oviedo, Oviedo, Spain
teranpedro@uniovi.es

Abstract. The relationship between possibility measures and Akian's semiring-valued idempotent probabilities is studied. The semiring operations in $[0,1]$ for which both theories overlap are shown to be the maximum and an arbitrary 1-Lipschitz triangular norm. Sufficient conditions are then given for the notion of convergence in law to be independent of the choice of a specific triangular norm.

Keywords: Convergence in law · Decision variable · Idempotent semiring · Portmanteau lemma · Possibility measure · Weak convergence

1 Introduction and Preliminaries

Akian [1] presented the basics of an 'alternate' probability theory where an idempotent semiring $(D, \oplus, \otimes)$ replaces the usual semiring $([0,\infty), +, \cdot)$. This is an abstract generalization of earlier work in the specific semiring $((-\infty,\infty], \min, +)$, see [2,3] and references therein. Her work was quickly noticed by the possibility community [6] but, to the best of my knowledge, the overlap area with [1] was never analyzed.

The existence of some overlap is obvious, as one easily figures that in the semiring $([0,1], \max, \cdot)$ the notions of idempotent probability, independent decision variables, convergence in cost and the Maslov integral will match possibility, product-related possibilistic variables, convergence in necessity and the Shilkret integral. Moreover, part of Akian's framework (e.g., her approach to L^p-like spaces) has not been developed independently by the possibility community. Hence a clarification of the connections is a useful addition to the possibility toolkit. Our own motivation comes from the study of limit theorems for possibilistic variables [14–16].

A *triangular norm* $\otimes$ (or *t-norm* for short) is an operation in $[0,1]$ which is associative, commutative, increasing in each argument, and has neutral element 1. By these properties, the meaning of an iterated operation $\bigotimes_{i=1}^{n} x_i$ is univocal. It will be essential that, if $\otimes$ is continuous, commutativity is implied by the other properties, a result which goes back to Faucett [7, Lemma 5]. A t-norm is called *1-Lipschitz* if $|(x \otimes y) - (x \otimes z)| \leq |y - z|$ for all $x, y, z \in [0,1]$ (this is equivalent

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 509–517, 2024.
https://doi.org/10.1007/978-3-031-65993-5_62

to $\otimes$ having Lipschitz constant 1 as a function on $[0,1]^2$ with the distance given by the sum norm).

Given a set Ω, a *possibility measure* is a set function $\Pi : \mathcal{P}(\Omega) \to [0,1]$ such that $\Pi(\emptyset) = 0$, $\Pi(\Omega) = 1$ and $\Pi(\bigcup_{i\in I} A_i) = \sup_{i\in I} \Pi(A_i)$ for every index set I[1]. From the definition, the function $\pi : \omega \in \Omega \to \pi(\omega) = \Pi(\{\omega\})$ satisfies $\Pi(A) = \sup_{\omega\in A} \pi(\omega)$. It will be called a *density function* (also *possibility distribution* and *possibility profile* in the literature).

A *possibilistic variable* (also called *fuzzy variable*) is a mapping $X : \Omega \to \mathbb{R}$. The *distribution* of X is the possibility measure $\Pi_X : A \mapsto \Pi(X \in A) = \Pi(X^{-1}(A))$. The density function π_X of Π_X will be called the density function of X. Accordingly, $\Pi_X(A) = \sup_{x\in A} \pi_X(x)$ for all $A \subseteq \mathbb{R}$.

A family $\{X_i\}_{i\in I}$ of possibilistic variables is $\otimes$ *-related* if, for any finite subset $\{i_j\}_{j=1}^n$ of distinct elements of I and all $A_1, \ldots, A_n \subseteq \mathbb{R}$,

$$\Pi(\bigcap_{j=1}^{n} \{X_{i_j} \in A_j\}) = \bigotimes_{j=1}^{n} \Pi(X_{i_j} \in A_j). \tag{1}$$

The indicator function of a set A will be denoted by I_A. A function $f : \mathbb{R} \to \mathbb{R}$ is *upper semicontinuous* if $x_n \to x$ implies $\limsup_n f(x_n) \leq f(x)$ or, equivalently, sets of the form $\{f \geq t\}$ are closed for $t \in \mathbb{R}$. Let $\mathcal{U}(\mathbb{R})$ denote the set of all (necessarily upper semicontinuous) functions $u : \mathbb{R} \to [0,1]$ such that $\{u \geq t\}$ is non-empty and compact for all $t \in (0,1]$. Functions in $\mathcal{U}(\mathbb{R})$ attain their supremum on any non-empty closed set.

A sequence of upper semicontinuous functions *hypo-converges* $f_n \to f$ if, for all $x \in \mathbb{R}$, one has $f(x) \geq \limsup_n f_n(x_n)$ whenever $x_n \to x$, and $f(x) \leq \liminf_n f_n(x_n)$ for some $x_n \to x$. This can equivalently be defined in terms of convergence of the *hypograph* sets $\{(x,t) \mid f_n(x) \geq t\}$ as in [18], see [10, pp. 158–160] for more details.

2 Akian Semirings in $[0,1]$

We begin by defining *Akian semirings*, i.e., those which satisfy all the assumptions collected in [1, Sect. 1]. The aim of this section is to give a simple characterization of the Akian semirings for which the unit interval with its usual structure is the range of a semiring-valued probability.

Definition 1. An *Akian semiring* $(D, \oplus, \otimes, \preceq, d)$ is a set D endowed with operations $\oplus, \otimes$, a partial order $\preceq$ and a metric d for which the following properties hold. (All properties below which are not explicitly quantified are being required for all choices of elements in D.)

(SR) $(D, \oplus, \otimes)$ is a *semiring*, i.e.,

[1] Due to space restrictions, possibility measures and idempotent probabilities defined on more general classes of sets will not be discussed. In both settings, one can pass from the general case to the case of $\mathcal{P}(\Omega)$.

(Mo) $\oplus$ and $\otimes$ are associative with respective neutral elements $0_D, 1_D$ (monoid properties), and $\oplus$ is commutative.
(An) $x \otimes 0_D = 0_D \otimes x = 0_D$ (annihilator).
(Di) $(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$ and $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$ (distributivity).
(ID) $\oplus$ is *idempotent*, i.e., $x \oplus x = x$.
(PO) $\preceq$ is the partial order induced by $\oplus$ in the sense that $x \preceq y \Leftrightarrow x \oplus y = y$.
(BCL) $(D, \preceq)$ is a lattice and is *boundedly complete* (every set with an upper bound has a least upper bound); the lattice operations will be denoted by $\vee$ and $\wedge$ (under the former assumptions, D is necessarily a sup-semilattice with $\vee = \oplus$.)
(CDS) $\otimes$ is completely distributive with respect to taking suprema: for any index set I, we have $x \otimes (\bigvee_{i \in I} y_i) = \bigvee_{i \in I} (x \otimes y_i)$ whenever the supremum $\bigvee_{i \in I} y_i$ exists.
(D1) $d(x \vee x', y \vee y') \leq \max\{d(x, y), d(x', y')\}$.
(D2) There exists a mapping $\Delta : D \to [0, \infty)$ with the following properties:

(δ1) $\Delta(0_D) = 0$, $\Delta(1_D) = 1$, and $\Delta(x \vee y) = \max\{\Delta(x), \Delta(y)\}$.
(δ2) $d(a \otimes x, a \otimes y) \leq \Delta(a) \cdot d(x, y)$ and $d(x \otimes a, y \otimes a) \leq \Delta(a) \cdot d(x, y)$.
(δ3) $\Delta(x \otimes y) \leq \Delta(x) \cdot \Delta(y)$.
(δ4) Δ maps d-bounded sets of D to bounded sets of $[0, \infty)$; please notice the difference between boundedness in the metric sense and in the lattice sense.

(D3) $d(x_2, y) \leq \max\{d(x_1, y), d(x_3, y)\}$ whenever $x_1 \preceq x_2 \preceq x_3$.
(D4) $d(x \wedge x', y \wedge y') \leq \max\{d(x, y), d(x', y')\}$.
(D5) If $\{x_i\}_{i \in I}$ is a net (nets are a generalization of sequences, see [19, Definition 11.2]) in D which is increasing (resp. decreasing) for $\preceq$ and there exists $x = \bigvee_{i \in I} x_i$ (resp. $x = \bigwedge_{i \in I} x_i$) then $d(x_i, x) \to 0$.

Remark 1. In this definition, one starts with an idempotent operation $\oplus$, constructs $\preceq$ and obtains $\oplus = \vee$. One could also begin directly with a lattice and then the join operation $\vee$ is of course idempotent. In that sense, the symbol $\oplus$ is unnecessary and the reader may check that I have replaced some $\oplus$ by $\vee$ in the belief that this makes the rationale behind some properties easier to grasp. However, the point of the notation $\oplus, \otimes$ is to make it apparent that probabilistic notions, results and proofs involving sums and products translate into analogs with the semiring operations $\oplus$ and $\otimes$.

Remark 2. In [1, Sect. 2], the assumption that D, with its reverse order, is a continuous lattice is made in order to define the analog of the Lebesgue integral (see next section). Since $([0, 1], \geq)$ *is* a continuous lattice and this assumption is not necessary for Theorem 1, it is not included in the definition.

Let D be an Akian semiring. An *idempotent probability* [1, p. 9] on $(\Omega, \mathcal{P}(\Omega))$ is a mapping $\mu : \mathcal{P}(\Omega) \to [0_D, 1_D]$ such that $\mu(\emptyset) = 0_D, \mu(\Omega) = 1_D, \mu(A \cup B) = \mu(A) \oplus \mu(B), A_n \nearrow A \Rightarrow \mu(A_n) \to \mu(A)$. If $\mu(A) = \sup_{\omega \in A} c(\omega)$ holds for a function c, it is called a *density* of μ.

In order to identify which Akian semiring structures in $[0,1]$ can work as the range of a possibility measure, $0_D = 0$ always holds but the assumption $1_D = 1$ must be added.

Theorem 1. *Let $[0,1]$ be endowed with its usual order $\leq$ and Euclidean metric d. Then the following are equivalent:*

(1) $([0,1], \oplus, \otimes, \leq, d)$ *is an Akian semiring with* $1_D = 1$,
(2) $\oplus$ *is the maximum and* $\otimes$ *is a 1-Lipschitz triangular norm.*

Proof. *(1) $\Rightarrow$ (2):* As stated in Definition 1, the join operation $\vee$ of the lattice equals $\oplus$. Since we are considering $[0,1]$ with its usual order, $x \vee y = \max\{x, y\}$, i.e., $\oplus$ must be the maximum.

From the distributive law $(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$ there follows

$$x \leq y \Rightarrow y = x \oplus y \Rightarrow y \otimes z = (x \oplus y) \otimes z \Rightarrow x \otimes z \leq y \otimes z.$$

Analogously, $x \leq y \Rightarrow z \otimes x \leq z \otimes y$. Thus, for $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$x_1 \otimes y_1 \leq x_2 \otimes y_1 \leq x_2 \otimes y_2$$

which proves $\otimes$ is increasing in each argument. Its associativity is included in (Mo). Moreover, also by (Mo), $D = [0,1]$ has a neutral element 1_D for $\otimes$, while by assumption $1_D = 1$.

To establish its 1-Lipschitzianity,

$$d(x \otimes y, x \otimes z) \leq \Delta(x) d(y,z) \leq \Delta(1) d(y,z) = d(y,z)$$

where one succesively uses $(\delta 2)$, the third part of $(\delta 1)$, and finally $1 = 1_D$ and the second part of $(\delta 1)$.

As commented in the Preliminaries, 1-Lipschitzianity implies joint continuity and, by Faucett's lemma, commutativity. Therefore $\otimes$ is a triangular norm.

(2) $\Rightarrow$ (1): As regards the metric's properties (D1, D3, D4) are well known while (D5) is, e.g., [5, Example O, p. 44]. To establish (D2), we take $\Delta(x) = I_{(0,1]}(x)$. Since $0_D = 0$ because $x \oplus 0 = x$ for all $x \in [0,1]$, Properties $(\delta 1, \delta 4)$ are immediately checked. As regards $(\delta 2)$, $0 \otimes x = x \otimes 0 = 0$ (for $0 \otimes x \leq 0 \otimes 1 = 0$ by the definition of a t-norm). That yields the case $a = 0$, while the case $a > 0$ follows from $\Delta(a) = 1$ and the 1-Lipschitzianity of $\otimes$. Similarly, $(\delta 3)$ is $0 \leq \Delta(x)\Delta(y)$ if $x \otimes y = 0$ and $\Delta(x \otimes y) \leq 1$ if $x \otimes y > 0$, both of which are clear.

Properties (ID), (PO) and (BCL) are clearly satisfied by $[0,1]$. As regards (SR), part (Mo) is easily checked, with $1_D = 1$ since $\otimes$ is a triangular norm. Part (An) was established above. Part (Di) follows from $\otimes$ being increasing as a bivariate function: since $\leq$ is a total order, assume without loss of generality $x \leq y$. Then $x \otimes z \leq y \otimes z$ and

$$(x \oplus y) \otimes z = y \otimes z = \max\{x \otimes z, y \otimes z\} = (x \otimes z) \oplus (y \otimes z).$$

The other distributive law is similar but also follows immediately from the commutativity of triangular norms.

As regards (CDS), since the usual order is being considered, we have $\bigvee_{i\in I} = \sup_{i\in I}$ for any index set I. Firstly, inequality $x \otimes (\bigvee_{i\in I} y_i) \geq \bigvee_{i\in I}(x \otimes y_i)$ holds: the left-hand side is an upper bound of each $x \otimes y_i$ (since $\otimes$ is increasing) and thus of their supremum. Secondly, there exists a sequence $\{y_{i_n}\}_n$ for which $y_{i_n} \to \sup_{i\in I} y_i$. Since $\otimes$ is continuous,

$$x \otimes (\bigvee_{i\in I} y_i) = \lim_n x \otimes y_{i_n}$$

where each term of the sequence is bounded above by $\bigvee_{i\in I}(x \otimes y_i)$, yielding the converse inequality $x \otimes (\bigvee_{i\in I} y_i) \leq \bigvee_{i\in I}(x \otimes y_i)$.

A *decision variable* [1, p. 13] is a mapping $X : \Omega \to \mathbb{R}$. Events $A, B \subseteq \mathbb{R}$ are called *independent* if $\mu(A \cap B) = \mu(A) \otimes \mu(B)$. From this definition, one can also define *independent decision variables* in a way formally identical to (1).

Corollary 1. *Let $([0,1], \oplus, \otimes, \leq, d)$ be an Akian semiring with the usual order, the Euclidean metric, and $1_D = 1$. A set function $\Pi : \mathcal{P}(\Omega) \to [0,1]$ is a possibility measure if and only if it is an idempotent probability with a density. A family of possibilistic variables $X_i : \Omega \to \mathbb{R}$ is $\otimes$-related if and only they are independent as decision variables.*

While $\otimes$ is fixed in Akian's framework as defining the independence property, different families of possibilistic variables can be $\otimes$-related by different triangular norms. Although the property that $\Pi(A \cap B) = \Pi(A) \otimes \Pi(B)$ is formally analogous to probabilistic independence, there is no guarantee that it admits the same semantics. In possibility theory, the case $\otimes = \min$ is interpreted as non-interactivity [20, p. 17] or unrelatedness [9, p. 99], while in Dempster–Shafer theory the case $\otimes = \cdot$ is interpreted as cognitive independence [11, p. 150].

3 Convergence in Law

Let $\gg$ be the *way below* relation [8, Definition 1.1] associated to the reverse order of D. In our context, $D = [0,1]$ and $\gg$ is just $>$. Akian defines the *Maslov integral* of X against an idempotent probability μ to be

$$I_\otimes[X;\mu] = \bigvee_{t\in D} t \otimes \mu(X \gg t).$$

We extend this definition to the case that $\otimes$ is a continuous (rather than 1-Lipschitz) t-norm. Following [1, Proposition 13 and p. 11], one shows that these Maslov integrals can alternatively be regarded as t-normed integrals [12] or as the global optima of certain functions.

Lemma 1. *Let X be a possibilistic variable with density function $\pi_X \in \mathcal{U}(\mathbb{R})$. Let $\otimes$ be a continuous triangular norm and let $f : \mathbb{R} \to [0,1]$. Then*

$$I_\otimes[f(X);\Pi] = I_\otimes[f;\Pi_X] = \sup_{t\in[0,1]} t \otimes \Pi(f(X) \geq t) = \sup(f \otimes \pi_X)(\mathbb{R}).$$

An important consequence due to the continuity and monotony of $\otimes$ is the fact that $I_\otimes[\sup_{i\in I} f_i(X);\Pi] = \sup_{i\in I} I_\otimes[f_i(X);\Pi]$ for any index set I.

A sequence $\{X_n\}_n$ of decision variables converges *in law* or *weakly* to X if $I_\otimes[f(X_n);\Pi] \to I_\otimes[f(X);\Pi]$ for all continuous $f : \mathbb{R} \to D = [0,1]$ (the requirement in [1] that f be bounded is void here.) Since this definition depends on the choice of $\otimes$, we will call it *weak convergence under* $I_\otimes$.

In possibility theory there is no *a priori* chosen $\otimes$, whence the fact that the definition of convergence in law depends on such a choice is inconvenient. Let us establish an analog of the portmanteau lemma from probability theory [4, Theorem 2.1, p. 16] which overcomes that situation under clear assumptions.

Theorem 2. *Let X_n, X be possibilistic variables with density functions in $\mathcal{U}(\mathbb{R})$. Let $\otimes$ be a continuous triangular norm. Then the following are equivalent:*

(i) $X_n \to X$ *weakly under* $I_\otimes$,
(ii) $\limsup_n \Pi(X_n \in F) \le \Pi(X \in F)$ *and* $\liminf_n \Pi(X_n \in G) \ge \Pi(X \in G)$ *for all closed F and open G.*

Proof. (i)⇒(ii): The argument in [4, p. 16] to establish $\limsup_n P(X_n \in F) \le P(X \in F)$ approximates the indicator I_F by a decreasing sequence of continuous functions with $I_F \le f \le I_{F_m}$ and makes $m \to \infty$ (where F_m is the set of all points at a distance from F at most m^{-1}). In order to replicate it with possibility measures, we just need to ensure that any decreasing sequence of closed sets $F_m \searrow F$ has $\Pi_X(F_m) \to \Pi_X(F)$. Since $\pi_X \in \mathcal{U}(\mathbb{R})$, there exist x_m such that $\pi_X(x_m) = \Pi_X(F_m)$. Reasoning by contradiction, assume the convergence fails. Then, for some $\varepsilon > 0$ the sequence $\{x_m\}_m$ is contained in the compact set $\{\pi_X \ge \Pi_X(F) + \varepsilon\}$, and it has a subsequence $\{x_{m'}\}_m$ converging to some $x \in \bigcap_m F_m = F$. Due to the upper semicontinuity, we reach a contradiction:

$$\Pi_X(F) \ge \pi_X(x) \ge \limsup_m \pi_X(x_{m'}) \ge \Pi_X(F) + \varepsilon.$$

Unlike in the probabilistic setting, the inequality for open sets is *not* equivalent to that for closed sets. But since each open G can have I_G approximated from below by continuous functions,

$$\begin{aligned}
&\liminf_n \Pi(X_n \in G) = \liminf_n I_\otimes[I_G;\Pi_{X_n}] = \liminf_n I_\otimes[\sup_{0\le g\le I_G, g \text{ cont.}} g;\Pi_{X_n}] \\
&= \liminf_n \sup_{0\le g\le I_G, g \text{ cont.}} I_\otimes[g;\Pi_{X_n}] \ge \sup_{0\le g\le I_G, g \text{ cont.}} \liminf_n I_\otimes[g;\Pi_{X_n}] \\
&= \sup_{0\le g\le I_G, g \text{ cont.}} I_\otimes[g;\Pi_X] = I_\otimes[I_G;\Pi_X] = \Pi(X \in G).
\end{aligned}$$

(ii)⇒(i): Since π_{X_n}, π_X are upper semicontinuous, in Vervaat's terminology [18] they are the *sup derivatives* of the *sup measures* obtained by restricting Π_{X_n}, Π_X to the family of open sets of $\mathbb{R}$. By [18, Remark 3.2.(b)], property (ii) amounts exactly to his *sup weak* convergence, which implies *sup vague* convergence $\Pi_{X_n} \to \Pi$ (obtained by replacing closed sets by compact sets in (ii))

and, by definition, the *sup vague* convergence $\pi_{X_n} \to \pi_X$ [18, Definition 3.3]. Moreover, the latter is equivalent to hypo-convergence $\pi_{X_n} \to \pi_X$ by [18, Theorem 7.2].

Now by [17, Table 1, Row 3, D⇔E], the mapping $u \mapsto \sup u(K)$ is upper semicontinuous (where u ranges over upper semicontinuous non-negative functions with the topology of sup vague convergence) and $u \mapsto \sup u(G)$ is lower semicontinuous. Take any arbitrary continuous function $f : \mathbb{R} \to [0,1]$. By the hypo-convergence, $\liminf_n \pi_{X_n}(x_n) \geq \pi_X(x)$ whenever $x_n \to x$, and $\limsup_n \pi_{X_n}(x_n) \leq \pi_X(x)$ for some $x_n \to x$. Since f and $\otimes$ are continuous, the functions $f \otimes \pi_n$ and $f \otimes \pi$ are upper semicontinuous. With the continuity and monotony of $\otimes$, one easily checks

$$\liminf_n (f \otimes \pi_{X_n})(x_n) \geq \lim_n f(x_n) \otimes \liminf_n \pi_{X_n}(x_n) \geq f(x) \otimes \pi_X(x) = (f \otimes \pi_X)(x)$$

for all $x_n \to x$ as well as $\limsup_n (f \otimes \pi_{X_n})(x_n) \leq (f \otimes \pi_X)(x)$ for some $x_n \to x$. That is, $f \otimes \pi_{X_n}$ hypo-converges (equivalently, sup vaguely converges) to $f \otimes \pi_X$. Accordingly,

$$\limsup_n \sup(f \otimes \pi_{X_n})(K) \leq \sup(f \otimes \pi_X)(K) \tag{2}$$

for all compact $K \subseteq \mathbb{R}$ and

$$\liminf_n \sup(f \otimes \pi_{X_n})(G) \geq \sup(f \otimes \pi_X)(G) \tag{3}$$

for all open $G \subseteq \mathbb{R}$.

Since the sequence $\{\Pi_{X_n}\}_n$ converges sup weakly, it is relatively compact in the topology of sup weak convergence, which is equivalent by [17, Table 1, Row 3, H] to the *equitightness* property that $\sup_n \Pi_{X_n}(K^c) \to 0$ along the directed family of all compact sets. There are two cases now.

Case 1: $f \otimes \pi_X$ is the null function. For any fixed $\varepsilon > 0$, using (2),

$$\limsup_n \sup(f \otimes \pi_{X_n})(\mathbb{R})$$

$$\leq \max\{\limsup_n \sup(f \otimes \pi_{X_n})(K), \limsup_n \sup(f \otimes \pi_{X_n})(K^c)\}$$

$$\leq \max\{0, \sup_n \sup \pi_{X_n}(K^c)\} = \sup_n \Pi_{X_n}(K^c) < \varepsilon$$

for some compact set K. Thus, by Lemma 1 and the arbitrariness of ε,

$$I_\otimes[f(X); \Pi] = \sup(f \otimes \pi_X)(\mathbb{R}) = 0$$

$$= \limsup_n \sup(f \otimes \pi_{X_n})(\mathbb{R}) = \limsup_n I_\otimes[f(X_n); \Pi],$$

yielding $I_\otimes[f(X_n); \Pi] \to 0 = I_\otimes[f(X); \Pi]$.

Case 2: $\alpha := \sup(f \otimes \pi_X)(\mathbb{R}) > 0$. The equitightness above is tantamount to the *tightness* in [13], which by [13, Proposition 8] is equivalent to the existence of some $u \in \mathcal{U}(\mathbb{R})$ such that $\pi_{X_n} \leq u$ for all n. Since $\otimes$ is a triangular norm, $f \otimes \pi_{X_n} \leq 1 \otimes \pi_{X_n} = \pi_{X_n}$ and $f \otimes \pi_X \leq \pi_X$. Thus all $f \otimes \pi_{X_n}$ and $f \otimes \pi_X$ are

dominated by $\max(u, \pi_X) \in \mathcal{U}(\mathbb{R})$. Since, by (3), $\liminf_n \sup(f \otimes \pi_{X_n})(\mathbb{R}) \geq \alpha > 0$, for all sufficiently large n we have $\sup(f \otimes \pi_{X_n})(\mathbb{R}) > \alpha/2$. Accordingly, the global maxima of $f \otimes \pi_{X_n}$ (for those n) and $f \otimes \pi_X$ are attained in the set $\{\max(u, \pi_X) \geq \alpha/2\}$, which is compact. Finally, apply inequalities (2) and (3) with $K = \{\max(u, \pi_X) \geq \alpha/2\}$ and $G = \mathbb{R}$ to obtain $\lim_n \sup(f \otimes \pi_{X_n})(\mathbb{R}) = \sup(f \otimes \pi_X)(\mathbb{R})$. By Lemma 1, that is exactly (i).

Corollary 2. *Let X_n, X be possibilistic variables with density functions in $\mathcal{U}(\mathbb{R})$, and let $\otimes$ be a continuous triangular norm. Then the fact whether $X_n \to X$ weakly under $I_\otimes$ or not, does not depend on the choice of $\otimes$. In particular, convergence in law $X_n \to X$ is a univocal notion across all possible Akian semiring structures in Theorem 1.*

Acknowledgments. Research in this paper was supported by Spain (PID2019-104486GB-I00) and the Principality of Asturias (SV-PA-21-AYUD/2021/50897).

References

1. Akian, M.: Theory of Cost Measures: Convergence of Decision Variables. Research Report no. 2611, Unité de Recherche INRIA Roquencourt (1995). Available at https://hal.science/inria-00074074v1
2. Akian, M., Quadrat, J.-P., Viot, M.: Bellman processes. In: Cohen, G., Quadrat, J.-P. (Eds.) 11th International Conference on Analysis and Optimization Systems, pp. 302–311. Springer, London (1994)
3. Akian, M., Quadrat, J.-P., Viot, M.: Duality between probability and optimization. In: Gunawardena, J. (Ed.) Idempotency, pp. 331–353. Cambridge University Press, Cambridge (1998)
4. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, New York (1999)
5. Brown, A., Pearcy, C.: Introduction to Operator Theory I. Springer, New York (2012)
6. Dubois, D., Prade, H.: Possibility theory: qualitative and quantitative aspects. In: Smets, P. (Ed.) Quantified Representation of Uncertainty And Imprecision, pp. 169–226. Kluwer, Dordrecht (1998)
7. Faucett, W.M.: Compact semigroups irreducibly connected between two idempotents. Proc. Amer. Math. Soc. **6**, 741–747 (1955)
8. Gierz, G., Hofmann, K.H., Keimel, K., Lawson, J.D., Mislove, M., Scott, D.S.: A Compendium of Continuous Lattices. Springer, Berlin (1980)
9. Nahmias, S.: Fuzzy variables. Fuzzy Sets Syst. **1**, 97–110 (1978)
10. Salinetti, G., Wets, R.J.-B.: Glivenko-Cantelli type theorems: an application of the convergence theory of stochastic suprema. Ann. Oper. Res. **30**, 157–168 (1991)
11. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
12. Suárez García, F., Gil Álvarez, P.: Two families of fuzzy integrals. Fuzzy Sets Syst. **18**, 67–81 (1986)
13. Terán, P.: On convergence in necessity and its laws of large numbers. In: Dubois, D., Lubiano, M.A., Prade, H., Gil, M.Á., Grzegorzewski, P., Hryniewicz (Eds.) Soft Methods for Handling Uncertainty and Imprecision, pp. 289–296. Springer, Berlin (2008)

14. Terán, P.: Law of large numbers for the possibilistic mean value. Fuzzy Sets Syst. **245**, 116–124 (2014)
15. Terán, P.: A Unified Approach to the Laws of Large Numbers for Possibility Measures in the Context of More General Uncertainty Theories. Submitted for publication
16. Terán, P., Pis Vigil, E.: Contrasting two laws of large numbers from possibility theory and imprecise probability. In: Destercke, S., Denoeux, T., Cuzzolin, F., Martin, A. (Eds.) Belief Functions: Theory and Applications, pp. 243–251. Springer, Cham (2018)
17. Vervaat, W.: Narrow and vague convergence of set functions. Statist. Probab. Lett. **6**, 295–298 (1988)
18. Vervaat, W.: Random upper semicontinuous functions and extremal processes. In: Vervaat, W., Holwerda, H. (Eds.) Probability and Lattices, pp. 1–56. CWI, Amsterdam (1997)
19. Willard, S.: General Topology. Dover, Mineola (2004)
20. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. **100**, 9–34 (1999)

Robust Sparse PCA with R

Valentin Todorov(✉)

United Nations Industrial development Organization (UNIDO), Vienna, Austria
valentin@todorov.at

Abstract. Numerous methods have been proposed for improving the interpretability of principal component analysis (PCA) results by setting to zero coefficients in the components and thus creating sparse solutions. Few of these methods are capable for identifying atypical observations in the data and providing sparse and robust to outliers solutions at the same time. And finally, from a practical point of view, the availability of software is of utmost importance for the use of these methods in data analysis. We compare three methods for sparse and robust PCA for which R implementations are available at the Comprehensive R Archive Network (CRAN) and illustrate them on real data examples. For selecting the tuning parameter in these methods an alternative criterion, a robust version of the Index of sparseness is proposed and demonstrated to provide very good results.

Keywords: PCA · Robust · Sparse · R

1 Introduction

Principal component analysis (PCA) [5] is one of the most popular methods for data analysis and has a wide range of applications in science, technology, financial services and other fields where large data sets are ubiquitous. It searches for directions (principal components, or PCs), along which the variance of the data is maximized. Finding a small number k of uncorrelated linear combinations of the original variables that capture most of the structure of the original data it presents a suitable method for data reduction of multivariate data. The components of the linear combinations, called loadings, are collected in the columns of the orthogonal loadings matrix which serves as a transformation matrix for projecting the data on the reduced PC space. This matrix is also the link between the original variables and the principal components and thus it could serve as a means for interpreting the PCs. However, PCA usually tends to provide PCs which are linear combinations of all the original variables (by giving them non-zero loadings). Regarding the interpretability of the results it would be very helpful to reduce not only the dimensionality but also the number of used variables (ideally to relate each PC to only a few variables). Loadings matrices with many values exactly equal to zero are called *sparse loadings matrices* and such matrices will direct the interpretation of a particular principal component to

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 518–525, 2024.
https://doi.org/10.1007/978-3-031-65993-5_63

only a small subset of variables, not considering all variables as in the classical PCA. Obtaining sparse loading matrices yields a *sparse PCA*, which is especially helpful for analyzing high dimensional data sets. It is not surprising that vast research effort was devoted to this issue and various proposals have been introduced in the literature. The first such proposal was *SCoTLASS* [6] which applied a lasso penalty on the loadings in a PCA optimization problem. Later Zou et al. [15] reformulated PCA as a regression problem and used the elastic net to obtain a sparse version—SPCA. A number of modifications and improvements were developed in the meantime, see [13] for a comprehensive review.

Another drawback of the classical PCA and its sparse extensions is that being based on sample variances they are very sensitive to the presence of even a few atypical observations in the data. The outliers could artificially increase the variance in an otherwise uninformative direction and this direction will be determined as a PC direction. To cope with the possible presence of outliers in the data a number of robust methods have been proposed (see [11] for an overview), however only a few of them can provide robust and sparse solutions at the same time. The robust sparse PCA method of Croux et al. [2] replaces the sample variance in the projection pursuit formulation of SCoTLASS by a robust scale estimator. The proposed by Hubert et al. [4] ROSPCA algorithm is an extension of the ROBPCA algorithm [3] which first identifies a set of regular observations and then uses SCoTLASS to obtain sparse loadings. A third approach is proposed by Wang and Van Aelst [14] which is based on least trimmed squares and they call it LTSSPCA. All three methods are implemented in R and are freely available on the Comprehensive R Archive Network (CRAN).

The rest of the paper is organized as follows. In Sect. 2 some background is presented and the three main methods are briefly outlined. Special attention is paid to the selection of the tuning parameters. Numerical study on real data applications is described in Sect. 3. Final remarks are given in Sect. 4.

2 The Methods and Their Implementation in R

Consider an $n \times p$ data matrix $\boldsymbol{X}$. Without loss of generality we can assume that the column means of $\boldsymbol{X}$ are all zeros. Note that in the context of robust PCA, the centering has to be done also in a robust way. We are looking for linear combinations $\boldsymbol{t}_j$ that result from a projection of the centered data on a direction $\boldsymbol{p}_j$,

$$\boldsymbol{t}_j = \boldsymbol{X}\boldsymbol{p}_j \tag{1}$$

such that

$$\boldsymbol{p}_j = \underset{\boldsymbol{p}}{\operatorname{argmax}} \operatorname{VAR}(\boldsymbol{X}\boldsymbol{p}) \tag{2}$$

subject to $\|\boldsymbol{p}_j\| = 1$ and $Cov(\boldsymbol{X}\boldsymbol{p}_j, \boldsymbol{X}\boldsymbol{p}_l) = 0$ for $l < j$ and $j = 1, \ldots, q$ with $q \leq \min(n, p)$. The solutions of these maximization problems are obtained by solving a Lagrangian problem, and the result is that the principal components of $\boldsymbol{X}$ are the linear combinations whose components $\boldsymbol{p}_j$ are given by the eigenvectors of the covariance matrix $Cov(\boldsymbol{X})$, and the variances are the corresponding eigenvalues

$l_j = \mathrm{VAR}(\boldsymbol{X}\boldsymbol{p}_j)$. Classical PCA is obtained if the sample covariance matrix $\boldsymbol{S}$ is used. The vectors $\boldsymbol{t}_j$ are collected as columns in the $n \times q$ *scores* matrix $\boldsymbol{T}$, and the vectors $\boldsymbol{p}_j$ as columns in the *loadings* matrix $\boldsymbol{P}$. The eigenvalues l_j are arranged in the diagonal of the $q \times q$ diagonal matrix $\boldsymbol{\Lambda}$.

The most straightforward way to robustify PCA is to replace $\boldsymbol{S}$ by a robust version like for example MCD (see [11]). Another approach to robust PCA uses *projection pursuit* (PP) and calculates directly the robust estimates of the eigenvalues and eigenvectors. Directions are sought for, which maximize the variance of the data projected onto them. The advantage of this approach is that the principal components can be computed sequentially, and that one can stop after q components have been extracted. Thus, this approach is appealing for high-dimensional data, in particular for problems with $p \gg n$. Using the sample variance in the maximization problem would lead to classical PCA, and robust scale estimators result in robust PCA. Suitable robust measures are the squared median absolute deviation (MAD) or the more efficient Q_n [9]. An algorithm suitable for high-dimensional data, was proposed by Croux et al. [1]. This method, denoted by RPCA-GRID is implemented in R in the package **rrcov**, available on CRAN.

An alternative approach is provided by the ROBPCA method ([3]) which combines ideas from projection pursuit and robust covariance estimation. Both approaches were extended to sparse versions, to be discussed below.

SRPCA [2]: Sparseness is introduced by adding an L_1 constraint in the definition (2) which yields

$$\boldsymbol{p}_j = \underset{\boldsymbol{p}}{\mathrm{argmax}}\, \mathrm{VAR}(\boldsymbol{X}\boldsymbol{p}) - \lambda_j \|\boldsymbol{p}\|_1 \tag{3}$$

where λ_j is a tuning parameter. To have a similar degree of sparsity over the different principal components [2] reformulate Eq. (3) and remain with only one tuning parameter to select. Choosing VAR as the sample variance results in the SCOTLASS method, which is sparse but not robust. SRPCA and SCOTLASS are implemented in R in the packages **pcaPP** and **rrcovHD**, both available on CRAN.

ROSPCA [4]: The second method to consider, which was introduced by Hubert et al. [4], is based on the ROBPCA algorithm and can be viewed as a two-step procedure. It starts by computing the Stahel-Donoho outlyingness of each observation to find a subset of regular observations and then uses SCoTLASS to generate sparse loadings from this subset. This method is implemented in R in the package **rospca** available on CRAN.

LTSSPCA [14]: A third method, introduced by Wang and Van Aelst [14], follows a completely different approach and is based on least trimmed squares. It adds sparseness to the method proposed by Maronna [7] by modifying the objective function to include a penalty term and the sparse PCs are estimated sequentially using the projection deflation method. Finally, a reweighing step is conducted by removing both score and orthogonal outliers and re-estimating the sparse PCs from the remaining set of regular observations. This method is implemented in R in the package **ltsspca** available on CRAN.

Selection of the tuning parameter λ. All sparse PCA methods use some kind of hyper parameter which controls the balance between explained variance and degree of sparsity. In SCOTLASS, SRPCA and ROSPCA this is λ which controls the L_1 penalty - the larger the parameter the more of the loadings will become exactly zero. Unconstrained PCA is obtained with $\lambda = 0$. Croux et al. [2] proposed to use a BIC-like criterion which looks at the ratio of residual variances and the degree of sparsity of the loadings matrix. If $\boldsymbol{A}_q^s$ and $\boldsymbol{A}_q^u$ are the loading matrices from sparse and unconstrained PCA respectively

$$BIC_{CFF}(\lambda) = \frac{\text{VAR}(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_q^s(\boldsymbol{A}_q^s)^\top)}{\text{VAR}(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_q^u(\boldsymbol{A}_q^u)^\top)} + df(\lambda)\frac{log(n)}{n} \tag{4}$$

where $\text{VAR}(\boldsymbol{Y}) = \sum_{j=1}^{p} Q_n^2(Y_j)$ is the total robust variance of a matrix Y (replacing Q_n^2 by the sample variance gives the non-robust version of the criterion) and $df(\lambda)$ denotes the cardinality (that is the number of non zero loadings) of $\boldsymbol{A}_q^s$. A function for optimizing the BIC_{CFF} criterion and graphical tools are provided in the R package **pcaPP**. Hubert et al. [4] propose for ROSPCA a more traditional version of a BIC criterion given by

$$BIC_{HUB}(\lambda) = ln\left(\frac{1}{h_1 p}\sum_{i=1}^{h_1} OD_{(i)}^2(\lambda)\right) + df(\lambda)\frac{log(h_1 p)}{h_1 p} \tag{5}$$

where $OD_{(i)}^2$ is the i-th smallest orthogonal distance for the model when using λ as the sparsity parameter and h_1 is the cardinality of the set of "good" observations H_1 identified in the first stage of the ROSPCA algorithm. Function for optimizing the BIC_{HUB} criterion and graphical tools are provided in the R package **rospca**.

Alternatively to these two criteria we propose a robust version of the *Index of sparseness* introduced by Trendafilov [13] which gives very good results:

$$IS(\lambda) = PEV^s \times PEV^u \times \frac{df(\lambda)}{pq} \tag{6}$$

where PEV^s and PEV^u are the percentages of explained variance by the sparse and the unconstrained model respectively and $df(\lambda)$ is as defined above. The percentage of explained variance is calculated as $PEV = \text{VAR}(\boldsymbol{X}\boldsymbol{A}_q)/\text{VAR}(\boldsymbol{X})$ with the loading matrix $\boldsymbol{A}_q$ of the sparse or the unconstrained model. In all these three criteria the first part measures the quality of the fit whereas the last term penalizes for model complexity thus achieving a trade-off between accuracy and sparsity.

LTSSPCA also makes use of a tuning parameter, but the algorithm inherently selects this parameter and the user has no direct control over it. As a result it is not possible to choose ad-hoc degree of sparseness with LTSSPCA but on the other hand the built in hyperparameter selection procedure is much faster than this process in SRPCA and ROSPCA.

3 Example: Automatic Vehicle Recognition Data

The automatic vehicle recognition data set was used to study methods for automatic vehicle recognition [10]. This data set from the Turing Institute, Glasgow, Scotland, contains measures of shape features extracted from vehicle silhouettes. The images were acquired by a camera looking downward at the model vehicle from a fixed angle of elevation. The original data set contains 946 records of four model vehicles: a double decker bus, Cheverolet van, Saab 9000 and an Opel Manta 400 but for our example we will use only the class of the buses, so each of the 218 rows corresponds to a view of a bus silhouette, and contains 18 attributes of the image. The data are available in the R package **rrcov** as a data set `bus`. For another use of this data set see [8], page 213, Example 6.3. The variance of the columns measured by the robust measure Q_n^2 vary from 0 (for variables `V6` and `V9`) to 39. Therefore we remove the variables `V6` and `V9` from the analysis and divide each variable by the corresponding Q_n.

To select the number of components to keep we look at the cumulative percent variance (CPV) of the non-constrained models as suggested in [2] and in [4]. For this purpose we compute a large number of components and present the CPV in screeplots. The screeplots for the standard and robust non-sparse PCA are shown in Fig. 1. The first four classical PCs explain more than 96% of the total variance and the first four robust PCs explain more than 92%, therefore we decide to retain four components in all cases.

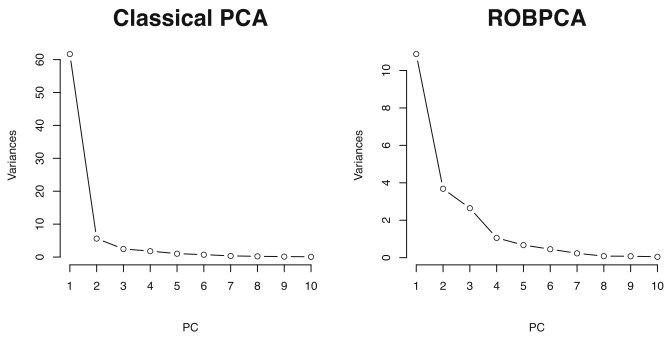


Fig. 1. Scree plots for standard and robust PCA (non-sparse) for the `bus` data set.

Next we need to choose the degree of sparseness which is controlled by the regularization parameter λ in SCOTLASS, SRPCA and ROSPCA. Since the sparse PCs have to provide a good trade-off between sparseness and achieved percentage of explained variance we can proceed similarly as in the selection of the number of principal components with the scree plot—we can compute the sparse PCA for many different values of λ and plot the percent of explained variance against λ or use one of the proposed information criteria. For SCOTLASS

the BIC_{CFF} criterion of [2] selects $\lambda = 8.4$ which excludes completely 12 variables and achieves full sparseness (one non-zero loading per PC) which results in 73% explained variance. The BIC_{HUB} of [4] selects $\lambda = 1.8$ which explains 88% of the variance but keeping 13 variables out of 16 in the PC space is not sparse enough. The index of sparseness (`IS`) selects $\lambda = 4.09$ which explains 83% of the variance excluding half of the variables and keeping three of the PCs with just a single non-zero loading.

The BIC_{HUB} criterion selects for ROSPCA $\lambda = 0.1$ which does not produce a sparse solution (only two zero loadings) while `IS` selects $\lambda = 2$ which excludes completely two variables and three of the PCs have only one non-zero loading each. The explained variance is 72%. It should be noted that grid for searching for λ was changed from the interval $[0, 2]$ to $[0, 10]$.

For SRPCA the BIC_{CFF} criterion selects $\lambda = 56.17$ which results in full sparseness and only 30% explained variance. The parameter selection tool of ROSPCA goes to the other extreme and selects $\lambda = 0.02$ which does not produce a sparse solution at all. The `IS` selects $\lambda = 5.46$ which explains 73% of the variance. Three of the four PCs are almost completely sparse, only on the leading PC remain 11 out of 16 non-zero loadings.

LTSSPCA also uses a BIC criterion but it is built in the estimation procedure and the user has not much control over the selection of the parameter λ. In our example it runs automatically and produces a solution which explains 79% of the variance but it is not sparse enough. All 16 variables are kept in the PC space and only one PC has just a single non-zero loading.

Using the fitted models we can produce outlier maps displaying the score distance and orthogonal distance of the observations. In order to make the results comparable between the different methods we normalize the distances by dividing each of them by the corresponding cutoff value. LTSSPCA has its own method for computing the score distances but this makes the output very different from the other methods, therefore we use the same definition of score distances. The results are shown in Fig. 2. There are two observations with very large values in variable V6 (MAX.LENGTH ASPECT RATIO), marked in blue, which are easily identified already by the classical methods. Interestingly, both ROSPCA and SRPCA find them as orthogonal outliers but not as bad leverage points. The observations marked in red have relatively large values in variable V7 (SCATTER RATIO) and are easily identified as bad leverage points by the robust methods. Only SRPCA does not find some of them. While the classical PCA does not identify any of these outliers, SCOTLASS marks half of them as orthogonal outliers.

Next we look at the sparsity of each of the considered methods on the first four principle components. ROBPCA and RPCA-GRID have almost no zero loadings. The greatest sparsity is attained by SCOTLASS keeping only 8 variables out of 16 in the PC space. The sparsity of the robust methods is quite similar, ROSPCA excludes two variables and SRPCA excludes one variable. By choosing larger λ in an ad-hoc way one could further increase the sparsity of ROSPCA and SRPCA.

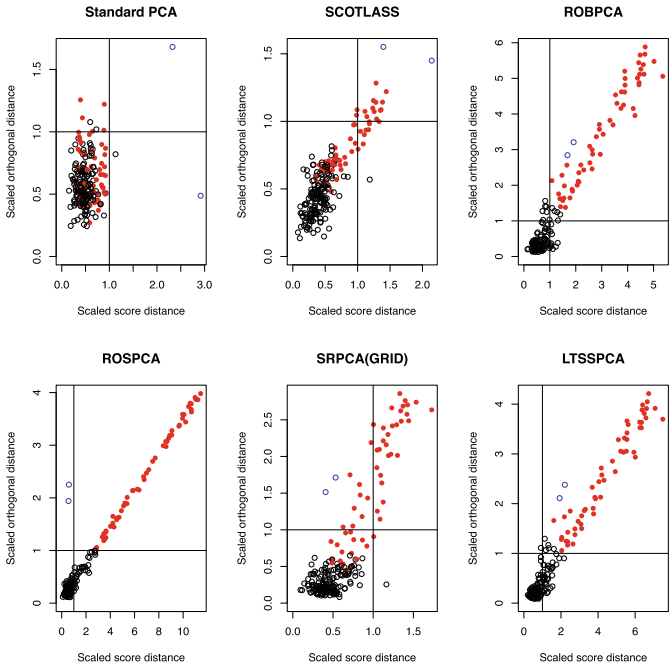


Fig. 2. Scaled outlier maps of standard PCA, SCOTLASS (with $\lambda = 4.09$), ROBPCA (non-sparse), ROSPCA (with $\lambda = 2.0$), SRPCA (with $\lambda = 5.46$) and LTSSPCA on the `bus` data.

Computing the last principal angle between the subspaces generated by the columns of the loadings, as this was done in [3,12], we can compare the obtained loadings across the methods. The classical PCA and SCOTLASS are very different from all of the methods with angles close to 1. The greatest similarity is shown between LTSSPCA and ROBPCA while the robust sparse methods ROSPCA, SRPCA and LTSSPCA are relatively different with angles larger than 0.5.

4 Summary and Conclusions

We compare three methods for sparse and robust PCA for which R implementations are available at the Comprehensive R Archive Network (CRAN) and illustrate them on a real data example. Special attention is paid to the selection of the tuning parameters. We propose a robust version of the *index of sparseness* [13] which seems to perform better than the originally proposed BIC-like criteria.

The example presented in Sect. 3 as well as many other examples that we have studied (not presented here due to space limitation) demonstrate that all three methods successfully present the data in lower dimensional space achieving sparseness of the loadings and identifying atypical observations present in

the data. LTSSPCA is fastest, especially in selecting the tuning parameter but not always produces sparse enough solutions. SRPCA, when tuned to higher sparseness could miss some of the outliers.

References

1. Croux, C., Filzmoser, P., Oliveira, M.: Algorithms for projection-pursuit robust principal component analysis. Chemom. Intell. Lab. Syst. **87**(218), 218–225 (2007)
2. Croux, C., Filzmoser, P., Fritz, H.: Robust sparse principal component analysis. Technometrics **55**(2), 202–214 (2013)
3. Hubert, M., Rousseeuw, P.J., Vanden Branden, K.: ROBPCA: a new approach to robust principal components analysis. Technometrics **47**, 64–79 (2005)
4. Hubert, M., Reynkens, T., Schmitt, E., Verdonck, T.: Sparse PCA for high-dimensional data with outliers. Technometrics **58**(4), 424–434 (2016)
5. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (2002)
6. Jolliffe, I.T., Trendafilov, N.T., Uddin, M.: A modified principal component technique based on the lasso. J. Comput. Graph. Stat. **12**, 531–547 (2003)
7. Maronna, R.A.: Principal components and orthogonal regression based on robust scales. Technometrics **47**, 264–273 (2005)
8. Maronna, R.A., Martin, D., Yohai, V.: Robust Statistics: Theory and Methods. Wiley, New York (2006)
9. Rousseeuw, P.J., Croux, C.: Alternatives to the median absolute deviation. J. Amer. Statist. Assoc. **88**(424), 1273–1283 (1993)
10. Siebert, J.P.: Vehicle Recognition Using Rule Based Methods. Turing Institute Research Memorandum TIRM-87-018 (1987)
11. Todorov, V., Filzmoser, P.: An object oriented framework for robust multivariate analysis. J. Stat. Softw. **32**(3), 1–47 (2009)
12. Todorov, V., Filzmoser, P.: Comparing classical and robust sparse PCA. In: Kruse, R., Berthold, M., Moewes, C., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (Eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis. Advances in Intelligent Systems and Computing, vol. 190, pp. 283–291. Springer, Berlin; New York (2013)
13. Trendafilov, N.: From simple structure to sparse components: a review. Comput. Stat. **29**, 431–454 (2014)
14. Wang, Y., Van Aelst, S.: Sparse principal component analysis based on least trimmed squares. Technometrics **62**(4), 473–485 (2020)
15. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. **15**, 265–286 (2006)

Dependent Censoring Based on Copulas

Ingrid Van Keilegom(✉)

ORSTAT, KU Leuven, Leuven, Belgium
ingrid.vankeilegom@kuleuven.be

Abstract. The paper considers survival models in which the survival time and the censoring time are stochastically dependent, which is referred to as dependent censoring. The non-identifiability of a fully non-parametric dependent censoring model leads to challenging problems. A common approach to handle this dependence is based on copulas. To overcome the non-identifiability of the model, the copula can be considered fully known. This is however a heavy assumption in practice, since the strength of the dependence is rarely known. Hence, it results in estimators that can be used for sensitivity analyses but rarely for point estimation of unknown quantities. Recently, a new approach to handle dependent censoring has been proposed, in which the copula is not fully known. The marginal distributions of the survival and censoring time can be modelled parametrically or semiparametrically. The paper describes the literature on these two streams of copula based models.

Keywords: Copulas · Dependent censoring · Identifiability · Cox regression

1 Introduction

In survival analysis we are interested in studying the time until an event of interest happens. This event can e.g. be the time until death for patients suffering from a certain disease, the time until an unemployed person finds a new job, or the time until a person who is waiting for an organ transplant receives a new organ. In this kind of situations, it often happens that the event of interest cannot be observed because of the presence of random right censoring. This means that instead of observing the event time of interest (denoted by T), we observe $Y = \min(T, C)$ and $\Delta = I(T \leq C)$, where C is a random variable, called the censoring time. So when C is smaller than T, we only observe a lower bound for the event time of interest. This can happen, for instance, when the event of interest did not happen yet at the end of the study, or when other events happen before the event of interest (like e.g. dying of another disease, or being lost to follow up because of various reasons).

A crucial and fundamental assumption that is made in most papers in the literature on survival analysis is the assumption of independent censoring. With that we mean that the random variables T and C are stochastically independent.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 526–531, 2024.
https://doi.org/10.1007/978-3-031-65993-5_64

The reason why this assumption is commonly made, is that in general we observe either T or C, but not both. This means that the relation between T and C is not identifiable in general, and therefore the bivariate distribution of (T, C) (and so also the marginal distribution of T) is not identifiable in general (see [16] or [2] for more details). However, by assuming that T and C are independent, the model is identified. The independence assumption is the most natural assumption, and fortunately it holds true in many contexts. For instance, for patients who are censored because they are still alive at the end of the study, it is reasonable to assume that the time until the end of the study is unrelated to the event time. Similarly, for patients who die or who are lost to follow up for reasons unrelated to the event of interest, the independent censoring assumption will likely hold.

However, the independence assumption is not always met. For instance, in medical studies, patients may withdraw from the study because their health condition is deteriorating or because they are showing side effects which need alternative treatments (positive relation between T and C). It could also happen that their health condition has improved and therefore they no longer follow the treatment (negative relation between T and C). In unemployment studies, unemployed people with low chances on the job market could decide to go abroad to improve their chances, leading to censoring times that depend on the duration of unemployment. Finally, in transplant studies, often the length of time a patient has to wait before he gets transplanted (T) depends on his/her medical condition, so on his time to death (C).

Let us illustrate what the impact is of incorrectly assuming independent censoring. We focus on the estimation of the survival function $S(t) = P(T > t)$, which is a key quantity in survival analysis, since many other quantities like the mean, median, quantiles, and the mean residual life function can be deduced from it. If we blindly assume independent censoring, the estimation of the survival function $S(t)$ will in general be biased, and the bias will increase with the strength of the dependence. To show this we consider as an example the case where $(\log T, \log C)$ follow a bivariate standard normal distribution with correlation ρ. We generate an arbitrary sample of size $n = 400$ from the distribution of $(Y, \Delta) = (\min(T, C), I(T \leq C))$, and calculate the Kaplan-Meier ([10]) estimator, which is the nonparametric maximum likelihood estimator of the survival function $S(t)$ under independent random right censoring. In Fig. 1 the true survival function of a log-standard normal variable is compared with the Kaplan-Meier estimator for several values of ρ. It is clear that when $\rho = 0$ (i.e. independent censoring), the two curves are close together, whereas the more ρ deviates from 0, the more the two curves move away from each other. For positive values of ρ the Kaplan-Meier estimator lies above the true survival function, whereas for negative values of ρ it is the opposite.

2 Copula Based Approach: The Case of a Known Copula

Now that the problem of dependent censoring has been explained, we will focus on how dependent censoring can be modeled and estimated, taking however into

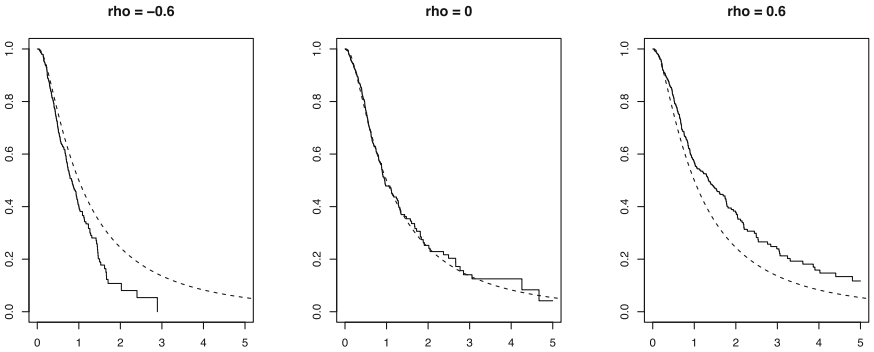


Fig. 1. Kaplan-Meier curves (solid curve) together with the true survival function (dashed curve) for $\rho = -0.6, 0$ and 0.6.

account that it cannot be identified in a fully nonparametric model, as explained earlier. A common approach in the literature is based on copula functions. A copula $\mathcal{C}$ is a bivariate distribution function defined on the unit square $[0,1] \times [0,1]$ with uniform margins. Copulas are a popular tool to model dependence between random variables, thanks to Sklar's theorem ([13]), which says that if the marginal distributions F and G are continuous, there exists a unique copula $\mathcal{C}$ such that

$$P(T \leq t, C \leq c) = \mathcal{C}(F(t), G(c)), \tag{1}$$

for all t, c, where $F(t) = P(T \leq t)$ and $G(c) = P(C \leq c)$ are the marginal distributions of T and C, respectively. Equivalently, we could also say that $P(T > t, C > c) = \tilde{\mathcal{C}}(\bar{F}(t), \bar{G}(c))$, where $\bar{F} = 1 - F, \bar{G} = 1 - G$ and $\tilde{\mathcal{C}}(u, v) = u + v + \mathcal{C}(1-u, 1-v) - 1$ is also a copula. Equation (1) shows that the copula $\mathcal{C}$ allows to disentangle the marginal distributions from the dependence between T and C, that is captured by the copula. Several sub-families of copula functions have been proposed, like the Archimedean copulas (including the well known Frank, Clayton and Gumbel copula), and the elliptical copulas (including the Gaussian copula). Most copula families depend on one parameter, say θ, which characterizes the strength of the dependence, and which is related to Kendall's tau τ via the formula:

$$\tau = 4 \int_0^1 \int_0^1 \mathcal{C}_\theta(u, v) d^2 \mathcal{C}_\theta(u, v) - 1.$$

More details about copulas can be found in the book by Nelsen ([11]).

Because of the above nice properties of copulas, it is very tempting to use copulas to model dependent censoring. This is exactly what Zheng and Klein ([18]) did: they obtained a nonparametric estimator of $F(t)$, called the copula-graphic estimator, that extends the Kaplan-Meier estimator to the case of dependent

censoring using the copula model in (1). The estimator is shown to be a nonparametric maximum likelihood estimator. However, their estimator requires the copula to be fully known, e.g. a Frank copula with known association parameter θ, which is quite restrictive in practice, since often one has no idea of the strength of the dependence. For that reason, their estimator is not very practical to obtain a point estimator of the marginal distribution of T, but it can be used for sensitivity analyses. Later, Rivest and Wells ([12]) obtained a closed form expression for the copula-graphic estimator in the special case where the copula belongs to the family of Archimedean copulas. For other copulas not belonging to this family the estimator needs to be computed iteratively.

The estimator of [12,18] has been extended to survival models with covariates by several authors. See e.g. [1] for fully nonparametric regression models, [17] for nonparametric regression quantiles, [8,9] for the Cox proportional hazards model, and [14,15] for a nonparametric location-scale model, among others. However, although the marginal distributions of T and C conditional on the covariates are non- or semiparametric in these papers, all these papers assume that the copula describing the relation between T and C is fully known, making these approaches less suitable in practice, unless the goal is to do sensitivity analyses.

3 Copula Based Approach: The Case of an Unknown Copula

In an attempt to weaken the conditions on the copula, Czado and Van Keilegom ([3]) proposed a copula model for dependent censoring without covariates, in which the copula is parametric but where the parameter, which describes the association between T and C, is unknown. There is however no free lunch, and the price to pay for this more flexible model on the copula, is that the marginal distributions of T and C are now parametric, whereas they were nonparametric in the papers by [12,18]. They developed high-level (or primitive) conditions under which their model is identified, and verified these conditions for specific parametric families for the copula and the marginal distributions. They showed that for most common parametric models, their conditions are satisfied.

The latter paper, which paved the road for more research on models for dependent censoring based on parametric copulas, was later generalized to parametric regression models by Deresa [5]. A relaxation of the parametric constraints on the conditional distributions of T and C (given covariates) is also possible. Deresa [6] showed under which conditions a semiparametric Cox proportional hazards model for T is identified, assuming that the copula and the distribution of C are parametric. As the assumption of a parametric distribution for C is still quite restrictive, [7] further generalized the model by assuming that both T and C follow a semiparametric transformation model. The latter model depends on a known transformation, which leads to the proportional hazards or proportional odds model for specific choices of this transformation. Finally, the case of dependent censoring in the presence of a cure fraction (i.e. a proportion of

subjects with infinite survival times) has been considered in [4], albeit in a fully parametric context without covariates.

4 Future Research

The approach to handle dependent censoring based on copulas that are not fully known, is still in full development. Many common models in survival analysis have not yet been extended to the case of dependent censoring based on unknown copulas. These include (among many others) the case of accelerated failure time models, location-scale models, quantile regression models, and outside the context of regression one could also consider problems like partial identification, goodness-of-fit testing, models with confounding factors, etc. Some of these extensions are currently in preparation. R code for all approaches described in the previous section (and beyond) is currently in preparation, and will allow to compute estimators in a user friendly and easily accessible way.

Acknowledgments. I. Van Keilegom gratefully acknowledges funding from the FWO and F.R.S. - FNRS (Excellence of Science programme, project ASTeRISK, grant no. 40007517), and from the FWO (senior research projects fundamental research, grant no. G047524N).

References

1. Braekers, R., Veraverbeke, N.: A copula-graphic estimator for the conditional survival function under dependent censoring. Can. J. Statist. **33**, 429–447 (2005)
2. Cox, D.: The analysis of exponentially distributed life-times with two types of failure. J. Roy. Statist. Soc. Ser. B **21**, 411–421 (1959)
3. Czado, C., Van Keilegom, I.: Dependent censoring based on parametric copulas. Biometrika **110**, 721–738 (2023)
4. Delhelle, M., Van Keilegom, I.: Copula based dependent censoring in cure models (under revision for TEST) (2024)
5. Deresa, N.W., Van Keilegom, I., Antonio, K.: Copula-based inference for bivariate survival data with left truncation and dependent censoring. Insur.: Math. Econ. **107**, 1–21 (2022)
6. Deresa, N.W., Van Keilegom, I.: Copula based Cox proportional hazards models for dependent censoring. J. Amer. Statist. Assoc. (to appear) (2024a)
7. Deresa, N.W., Van Keilegom, I.: Semiparametric transformation models for survival data with dependent censoring (submitted) (2024b)
8. Emura, T., Chen, Y.-H.: Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches. Springer, Singapore (2018)
9. Huang, X., Zhang, N.: Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. Biometrics **64**, 1090–1099 (2008)
10. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Statist. Assoc. **53**, 457–481 (1958)
11. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)

12. Rivest, L.P., Wells, M.T.: A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. J. Mult. Anal. **79**, 138–155 (2001)
13. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
14. Sujica, A., Van Keilegom, I.: Estimation of location and scale functionals in non-parametric regression under copula dependent censoring. Canad. J. Statist. **43**, 306–335 (2015)
15. Sujica, A., Van Keilegom, I.: The copula-graphic estimator in censored nonparametric location-scale regression models. Economet. Statist. **7**, 89–114 (2018)
16. Tsiatis, A. (1975): A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. **72**, 20–22 (2018)
17. Veraverbeke, N.: Regression quantiles under dependent censoring. Statistics **40**, 117–128 (2006)
18. Zheng, M., Klein, J.P.: Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika **82**, 127–138 (1995)

Wavelet Variance Based Robust Estimation of Composite Stochastic Models

Maria-Pia Victoria-Feser[1(✉)], Stephane Guerrier[2], and Roberto C. Molinari[3]

[1] Department of Statistical Sciences, University of Bologna, Bologna, Italy
maria.victoriafeser@unibo.it
[2] GSEM and Faculty of Sciences, University of Geneva, Geneva, Switzerland
[3] Department of Mathematics and Statistics, Auburn University, Alabama, USA

Abstract. Estimation and inference for complex parametric models related to signal and time-series analysis is still an extremely relevant area of research for many domains of application. In particular, with increasing data being collected and stored in different settings, there is a need for methods that provide these inferential solutions and that scale adequately with the large sample sizes. The availability of such methods is scarce and these can breakdown in particularly challenging circumstances also linked to the complexity of the models of interest. Another challenge that additionally constrains the availability of usable methods is the presence of contamination (outliers) in the signals which require even more tailored approaches to analyse them. In this context, the Robust Generalized Method of Wavelet Moments (RGMWM) was put forward to provide one of the few scalable and robust solution to perform inference for these complex stochastic models. Being based on a new bounded estimator of the Wavelet Variance and the previously proposed Generalized Method of Wavelet Moments, the RGMWM provides important solutions to model signals and time-series recorded in various domains going from engineering to finance, for the latter of which we also discuss an applied example.

Keywords: Wavelet variance · Scale-based analysis of variance · Large scale time series · Signal processing · Generalized method of wavelet moments · State-space models

1 Introduction

Despite the increased use and effectiveness of deep learning approaches (e.g., Recurrent Neural Networks) and other black-box models in fitting and predicting time series and other dependent processes, there is still an important need for parametric models with lower storage and processing costs as well as more interpretable outputs to characterize them. As a representative example, the field of navigation engineering is witnessing ongoing effort to achieve precise

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 532–539, 2024.
https://doi.org/10.1007/978-3-031-65993-5_65

calibration of low-cost inertial sensors that are increasingly used in applications involving unmanned vehicles and robotics. These efforts mainly consist in modeling the (deterministic and) stochastic measurement errors made by these sensors which are complex in nature since they represent an accumulation of different sources of error. For example, the stochastic errors often result from the sum of different latent stochastic processes which, in this work, we will refer to as *Latent* Time Series (LTS) models. In addition these errors are measured at high-frequency over a large amount of time to also detect long-term non-stationary features in these signals. Finally, due also to the sensitivity of these devices, in the majority of cases the recording of these signals is affected by different forms of contamination such as outliers and model deviations over brief intervals. For this purpose, a wide range of models and corresponding estimation methods exist, including the popular class of ARMA and state-space models for which likelihood-based approaches are commonly employed for estimation (see e.g., [7,21,24] for an overview). Nevertheless, the latter estimation approaches tend to become intractable or computationally unfeasible in many of the above-described circumstances (i.e. complex LTS models, large signals and contamination). Moreover, the existing robust approaches that aim to limit the impact of contamination (in the form of outliers) on the estimation process tend to be unusable even for simple models and moderate sample sizes (for motivating examples and robust inferential approaches, see e.g., [16,17]).

Considering the need for efficient approaches to extract information from time series and signals, wavelet decomposition represents a computationally efficient procedure to represent the information in signals across time-scales. Based on this, the Wavelet Variance (WV) is a quantity that is widely used to summarize this information into a reduced amount of quantities (see e.g., [20]). As a result of its usefulness, the latter has been used for analysis of variance, model building and prediction (see e.g., [21] for an overview) and has been extended to other fields of research for dependent processes (see e.g., [1,8,9,15,27,30] to mention a few). In line with this direction of research, [12] proposed a general inferential framework for time series models (including LTS) called the Generalized Method of Wavelet Moments (GMWM).[1] The GMWM optimizes a criterion based on a standardized distance between the sample WV estimates and the model-based WV. This alternative appraoch is straightforward to implement and often the only feasible estimation method available for complex models in different settings. It has been, so far, extended to model multivariate signals (see e.g., [28]), models with covariates [6] and to repeated signal recordings [2].

However, as highlighted in [18], the sample WV can be largely influenced by outlying observations, which then results in GMWM estimators that are not robust to the latter. To deal with the issue of outliers (contamination), traditional robust methods consists in placing estimation (and inference) in the framework of bounded Influence Function (IF) estimators that are obtained by bounding the corresponding estimating equations [13], such as the Maximum

[1] The method (with relative graphical tools) is implemented in an open source R package, available at www.github.com/SMAC-Group/gmwm; see also [3,5,22].

Likelihood Estimator (MLE) equations or indeed those of other efficient (non robust) estimators. In this perspective, robust methods for time series have been proposed in abundance over the past decades and a detailed overview of these can be found, for example, in [16], [17, Chap. 8]. These proposals are however not computationally feasible in many cases, especially in those dealing with complex models and large amounts of observations (see e.g., [11] and the refrences therein). To address this, in this short paper we provide an overview of the GMWM framework when incorporating bounded IF estimators, as proposed in [11]. Resulting in the Robust GMWM (RGMWM), this approach falls within the two-step robust framework, where the first step delivers a new robust estimator of the WV with appropriate asymptotic properties which, under standard conditions, are integrated within the second step where the GMWM matches these robust estimators with their model-implied counterparts.

2 Robust Wavelet Variance and the RGMWM

To introduce the WV, let us denote $(X_t \in \mathbb{R})$, $t = 1, \ldots, T$, as an *intrinsically* stationary time series that is either stationary or becomes so when a d^{th}-order backward difference is applied to it. Moreover, let

$$W_{j,t} := \sum_{l=0}^{L_j - 1} h_{j,l} X_{t-l} \,,$$

denote the wavelet coefficients that result from a wavelet decomposition of the time series, where $(h_{j,t})$ is a known wavelet filter of length L_j at scale τ_j (commonly, dyadic scale $\tau_j = 2^j$), for $j = 1, \ldots, J$ and $J < \log_2(T)$. Since the filter sequence $(h_{j,t})$ has specific properties, the wavelet coefficients can be seen as the result of a particular form of weighted moving-average taken over a number of observations proportional to L_j (which increases as j increases, see e.g., [21] for a general overview). Moreover, if $L_1 \geq \mathrm{d}$ and (X_t) is a dth-order stationary time series, the resulting wavelet coefficients will be stationary. Based on the (stationary) wavelet coefficients, for $j = 1, \ldots, J$, the WV is defined as

$$\nu_j^2 := \mathrm{Var}(W_{j,t}) \ > 0 \,,$$

i.e. the variance of the wavelet coefficients, which can be expressed in vector form as $\boldsymbol{\nu} := [\nu_j^2]_{j=1,\ldots,J}$. Throughout this work, we require $\nu_j^2 > 0$ and always consider $\nu_j = |\nu_j|$. This definition holds independently from a possible parametric model underlying the time series (X_t), but we assume that the latter is generated from the parametric family of models $F_{\boldsymbol{\theta}_0}$, with unknown parameter vector denoted by $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. In order to make the link between the WV and an assumed *stationary* parametric model $F_{\boldsymbol{\theta}}$ explicit, an alternative representation of this quantity is based on the Spectral Density Function (SDF) and is as follows:

$$\nu_j^2(\boldsymbol{\theta}) = \int_{-1/2}^{1/2} |H_j(f)|^2 S_{F_{\boldsymbol{\theta}}}(f) df \,, \tag{1}$$

with $S_{F_{\boldsymbol{\theta}}}(f)$ denoting the theoretical SDF and $H_j(f)$ being the Fourier transform. In practice, computing (1) is often difficult but the results of [29] provide the following result in terms of Autocovariance Functions (ACF):

$$\nu_j^2(\boldsymbol{\theta}) = \frac{2}{\tau_j^2}\left[\frac{\tau_i}{2}\left(1-\kappa\left(\boldsymbol{\theta},\tau_j/2\right)\right)+\sum_{i=1}^{\tau_j/2-1} i\left(2\kappa\left(\boldsymbol{\theta},\tau_j/2\right)-\kappa\left(\boldsymbol{\theta},i\right)-\kappa\left(\boldsymbol{\theta},\tau_j-i\right)\right)\right] \tag{2}$$

where $\tau_j = 2^j$ and $\kappa(\boldsymbol{\theta}, h)$ denotes the ACF, $\Sigma(\boldsymbol{\theta})$, of $\mathbf{X} = (X_1, \ldots, X_T)$, so that, for $h \in \mathbb{N}$ we have $\kappa(X, h) = \text{cov}(X_1, X_h) = \Sigma(\boldsymbol{\theta})_{1,h}$. Therefore, (2) provides a simple approach to compute the theoretical WV of nearly any (intrinsically) stationary parametric stochastic process (e.g., SARIMA and many state-space models). Indeed, the theoretical WV is a linear combination of the ACF $\kappa(\boldsymbol{\theta}, h)$, which composes the elements of $\Sigma(\boldsymbol{\theta})$. In some sense, the vector $\boldsymbol{\nu}(\boldsymbol{\theta}) = \left[\nu_j^2(\boldsymbol{\theta})\right]_{j=1,\ldots,J}$ of dimension $J < \log 2(n)$ summarizes the main characteristics of the matrix $\Sigma(\boldsymbol{\theta})$, by summarizing the ACF into few quantities ($J < \log_2(T)$).

An unbiased sample estimator of the WV, to whom $\boldsymbol{\nu}(\boldsymbol{\theta})$ is matched, has been proposed by [20] and is defined as

$$\tilde{\nu}_j^2 := \frac{1}{M_j}\sum_{t=1}^{M_j} W_{j,t}^2 \ , \tag{3}$$

where M_j is the length of the wavelet coefficient process $(W_{j,t})$ at scale τ_j. The theoretical properties of this estimator were further studied in [23] in which the conditions for its asymptotic properties are given. For a robust alternative, [11] propose to use "Huber's Proposal 2" which was presented in [14] and was aimed at the estimation of the scale parameter of the residuals in the linear regression framework. Let $r_{j,t} := W_{j,t}/\nu_j$ as the standardized wavelet coefficients with $\mathbb{E}[r_{j,t}] = 0$, [11] proposed

$$\hat{\nu}_j^2 := \underset{\nu_j^2 \in \mathbb{R}^+}{\text{argzero}} \left[\frac{1}{M_j}\sum_{t=1}^{M_j} \omega^2\left(r_{j,t};\nu_j^2,c\right)\, r_{j,t}^2 - a(\nu_j^2,c)\right], \tag{4}$$

where $\omega(\cdot)$ represents the robust weight function and $a(\nu_j^2, c)$ is a correction term to ensure Fisher consistency at the marginal distribution of the wavelet coefficients $(W_{j,t})$. For the robust weights, popular choices include Huber's weights and Tukey's Biweight weights (see e.g., [13]). The correction term is defined as

$$a(\nu_j^2, c) := \mathbb{E}\left[\omega^2\left(r_{j,t};\nu_j^2,c\right)\, r_{j,t}^2\right],$$

where the latter expectation is taken over the marginal distribution of $(W_{j,t})$ with variance ν_j^2. It can be noticed that, if the tuning constant $c \to \infty$, we have that $\omega(\cdot) \to 1$ and $a(\nu_j^2, c) \to 1$ such that $\hat{\nu}_j^2$ is the estimator in (3).

The properties of the proposed M-estimator of WV can be directly carried over to the GMWM framework (see [12]). Indeed, as suggested in [10], one can replace the standard estimator used in the GMWM with a robust estimator

which, in this case, is the proposed M-estimator in (4) allowing us to deliver the RGMWM defined as:

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^\top \boldsymbol{\Omega} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})), \tag{5}$$

where $\boldsymbol{\Omega}$ is any symmetric positive definite weighting matrix.

3 Application: Personal Saving Rates

In [11], the RGMWM framework is used for analysis on real data concerning personal saving rates. The data consist in the monthly seasonally adjusted Personal Saving Rates (PSR) from January 1959 to May 2015 provided by the Federal Reserve Bank of St. Louis.[2] The study of PSR is an essential part of the overall investigation on the health of national and international economies since, within more general economic models, PSR can greatly impact the funds available for investment which in turn determine the productive capacity of an economy. In this sense, [25] study the factors behind saving rates and investigate different models which, among others, are compared to the random-walk-plus-noise (local level) model (RWN). As opposed to the latter model, various time-varying models are proposed in the literature to explain precautionary PSR together with risk aversion in the light of different factors such as financial shocks or others (see, for example, [4,26]). Nevertheless, as emphasized in [19], modelling the time series with a stationary model, or a d^{th}-order non-stationary model such as an ARIMA, can be useful under many aspects such as, for example, to understand if a dynamic model is needed for forecasting and, if so, what kind of model is appropriate.

Using empirical representations of the data and a model selection criterion, [11] consider, as a suitable model, a random walk plus the sum of two AR(1) processes (i.e. ARMA(2,1)). The saving rates time series with different types of points indicating outliers identified through the weights of the proposed M-estimator is presented in Fig. 1 (taken from [11]). The outliers identified by the RGMWM can be interpreted in the light of the national and global economic and political events. Limiting ourselves to the major identified outliers, the first one corresponds to a rise in the precautionary savings in the aftermath of the OPEC oil crisis and the 1974 stock market crash. In the months following October 1987 we can see an instability in the PSR with a rise and sudden fall linked to the "Black Monday" stock market crash which added to the savings and loans crisis which lasted to the early 1990s. This period also saw an economic recession where a rise in the saving rates, highlighted by the presence of high outliers, led to a drop in aggregate demand and bankruptcies. Finally, the various financial crises of the 21^{st} century led to sudden and isolated rises in PSR as indicated again by the outliers.

[2] U.S. Bureau of Economic Analysis, Personal Saving Rate [PSAVERT], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PSAVERT.

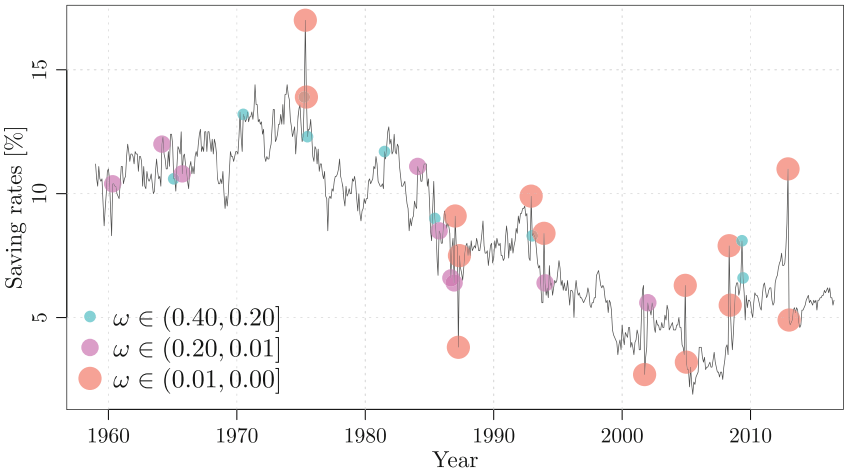


Fig. 1. Saving rates time series with different types of points indicating outliers identified through the weights of the proposed M-estimator.

4 Conclusion

In this short review, based on the works of [11,12] and the reference therein, we have presented the advantages of using a WV decomposition as a sufficient summary measure to estimate LTS models. Given that increasing amount of data are being collected and stored, the GMWV and RGMWM approaches adequately scale with large sample sizes and complex parametric models. The clear advantage is that a robust estimator for the model's parameters can be obtained in a computationally efficient manner, which is especially appropriate when storage and processing costs as well as more interpretable outputs to characterize the LTS are needed. An interactive short course can be found at https://gmwm.netlify.app/chapter4.

References

1. Abry, P., Didier, G.: Wavelet estimation for operator fractional Brownian motion. Bernoulli **24**(2), 895–928 (2018)
2. Bakalli, G., Cucci, D.A., Radi, A., El-Sheimy, N., Molinari, R., Scaillet, O., Guerrier, S.: Multi-signal approaches for repeated sampling schemes in inertial sensor calibration. IEEE Trans. Signal Process. **71**, 1103–1114 (2023)
3. Bakalli, G., Radi, A., Nassar, S., Guerrier, S., Zhang, Y., Molinari, R.: A two-step computationally efficient procedure for IMU classification and calibration. In: 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), pp. 534–540 (2018)
4. Brunnermeier, M.K., Nagel, S.: Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals. Am. Econ. Rev. **98**(3), 713–36 (2008)

5. Clausen, P., Skaloud, J., Molinari, R., Lee, J., Guerrier, S.: Use of a new online calibration platform with applications to inertial sensors. IEEE Aerosp. Electron. Syst. Mag. **33**(8), 30–36 (2018)
6. Cucci, D.A., Voirol, L., Kermarrec, G., Montillet, J.-P., Guerrier, S.: The generalized method of wavelet moments with exogenous inputs: a fast approach for the analysis of GNSS position time series. J. Geod. **97**, 14 (2023)
7. Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods. Oxford University Press (2012)
8. Foufoula-Georgiou, E., Kumar, P.: Wavelets in Geophysics. Wavelet Analysis and Its Applications, vol. 4. Academic Press (2014)
9. Gallegati, M.: A wavelet-based approach to test for financial market contagion. Comput. Stat. Data Anal. **56**(11), 3491–3497 (2012)
10. Guerrier, S., Molinari, R., Victoria-Feser, M.P.: Estimation of time series models via robust wavelet variance. Austrian J. Stat. **43**(4), 267–277 (2014)
11. Guerrier, S., Molinari, R., Victoria-Feser, M.P., Xu, H.: Robust two-step wavelet-based inference for time series models. J. Am. Stat. Assoc. **117**, 1996–2013 (2022)
12. Guerrier, S., Skaloud, J., Stebler, Y., Victoria-Feser, M.P.: Wavelet-variance-based estimation for composite stochastic processes. J. Am. Stat. Assoc. **108**(503), 1021–1030 (2013)
13. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York (1986)
14. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
15. Jia, X., An, H., Fang, W., Sun, X., Huang, X.: How do correlations of crude oil prices co-move? A grey correlation-based wavelet perspective. Energy Econ. **49**(C), 588–598 (2015)
16. Maronna, R.A., Martin, R.D., Yohai, V.J.: Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics. Wiley, Chichester, West Sussex, UK (2006)
17. Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M.: Robust statistics: theory and methods (with R). Wiley (2019)
18. Mondal, D., Percival, D.B.: M-estimation of wavelet variance. Ann. Inst. Stat. Math. **64**, 27–53 (2012)
19. Pankratz, A.: Forecasting with dynamic regression models. In: Wiley Series in Probability and Statistics, vol. 935. Wiley (2012)
20. Percival, D.B.: On estimation of the wavelet variance. Biometrika **82**(3), 619–631 (1995)
21. Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 4. Cambridge University Press (2006)
22. Radi, A., Bakalli, G., Guerrier, S., El-Sheimy, N., Sesay, A.B., Molinari, R.: A multisignal wavelet variance-based framework for inertial sensor stochastic error modeling. IEEE Trans. Instrum. Meas. **68**(12), 4924–4936 (2019)
23. Serroukh, A., Walden, A.T., Percival, D.B.: Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. J. Am. Stat. Assoc. **95**(449), 184–196 (2000)
24. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and its Applications. Springer Science and Business Media (2013)
25. Slacalek, J., Sommer, M.: What drives household saving? Unpublished note **10**, 95–107 (2012)
26. Videras, J., Wu, S.: The behavioral effects of financial shocks. Technical Report, Citeseer, Hamilton College (2004)

27. Xie, S., Krishnan, S.: Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis. Med. Biol. Eng. Comput. **51**, 49–60 (2013)
28. Xu, H., Guerrier, S., Molinari, R., Karemera, M.: Multivariate signal modelling with applications to inertial sensor calibration. IEEE Trans. Signal Process. **67**(19), 5143–5152 (2019)
29. Zhang, N.F.: Allan variance of time series models for measurement data. Metrologia **45**, 549–561 (2008)
30. Ziaja, A., Antoniadou, I., Barszcz. T., Staszewski, W.J., Worden, K.: Fault detection in rolling element bearings using wavelet-based variance analysis and novelty detection. J. Vib. Control **22**(2), 396–411 (2016)

Check for updates

Intersections of Randomly Translated Sets

Tommaso Visonà(✉)

Universität Bern, Institute of Mathematical Statistics and Actuarial Science, Bern, Switzerland
tommaso.visona@unibe.ch

Abstract. Let $\Xi_n = \{\xi_1, \ldots, \xi_n\}$ be a sample of n independent points distributed on a regular closed element K of the extended convex ring in $\mathbb{R}^d$ according to a probability measure μ on K, admitting a density function. We consider random sets generated from the intersection of the translations of K by elements of Ξ_n, as

$$X_n = \bigcap_{i=1}^{n} (K - \xi_i).$$

This work aims to show that scaled X_n as $n \to \infty$ converges in distribution to the zero cell of a Poisson hyperplane tessellation whose distribution is determined by the curvature measure of K and the behaviour of the density of μ near the boundary of K.

Keywords: Minkowski difference · Regular closed random sets · Zero cell of a poisson tesselation · Intersection of random sets

1 Introduction

A random closed set X in $\mathbb{R}^d$ is a measurable map from a probability space to the space $\mathcal{F}$ of closed sets in $\mathbb{R}^d$ endowed with the Borel σ-algebra generated by the Fell topology, see [7]. The subbase of the Fell topology consists of finite intersections of the sets $\{F \in \mathcal{F} : F \cap G \neq \emptyset\}$ and $\{F \in \mathcal{F} : F \cap L = \emptyset\}$ for all open G and compact L in $\mathbb{R}^d$.

It is known that the distribution of a random closed set is uniquely determined by its capacity functional defined as

$$T_X(L) := \mathbb{P}\{X \cap L \neq \emptyset\},$$

where L runs through the family of compact sets in $\mathbb{R}^d$. A sequence of random closed sets $(X_n)_{n\geq 1}$ in $\mathbb{R}^d$ converges in distribution to a random closed set X if the corresponding probability measures on $\mathcal{F}$ weakly converge. This is the case if and only if

$$T_{X_n}(L) \to T_X(L) \quad \text{as } n \to \infty$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 540–547, 2024.
https://doi.org/10.1007/978-3-031-65993-5_66

for each compact set L such that $T_X(L) = T_X(\mathrm{Ind}(L))$, see Theorem 1.7.7 from [7].

The capacity functional agrees well with the Minkowksi sum and the union operation. Many limit theorems for random sets are derived for these operations. On the contrary, the capacity functional does not work well with the intersection operation.

Only recently, [8] proved a limit theorem for random sets obtained as the intersection of unit Euclidean balls $x + B_1$ whose centres x form a Poisson point process of growing intensity λ on the same unit ball. It was shown that these random sets scaled by λ converge in distribution as $\lambda \to \infty$ to the zero cell of a Poisson hyperplane tessellation. It was also shown that the volumes of these intersection sets converge in distribution.

Many results from [8] follow from the studies of a similar model. It appears that by taking the ball hull of a sample $\Xi_n = \{\xi_1, \dots, \xi_n\}$ which consists of i.i.d. points uniformly distributed in the unit Euclidean ball. The ball hull

$$Q_n = \bigcap_{x \in \mathbb{R}^d, \Xi_n \subset x + B_1} (x + B_1)$$

is the intersection of all (translated) unit balls which contain the sample. It is easy to see that the set X_n of all $x \in \mathbb{R}^d$ such that $\Xi_n \subset x + B_1$ satisfies

$$X_n = \{x : \Xi_n \subset x + B_1\} = \bigcap_{i=1}^{n} (\xi_i + B_1).$$

This line of research was initiated in [1], where the the combinatorial structure of the ball polytope Q_n was explored. In particular, it was shown that the expectation of the number of faces of Q_n in dimension 2 converges to a nontrivial limit without taking any normalisation. This was generalized in [2], by replacing the ball with a ball polytope in $\mathbb{R}^2$.

The ball hull model has been extended in [6], where the unit ball was replaced by a general convex body in the space of arbitrary dimension and it was shown that the convergence of distribution of nX_n entails the convergence of the combinatorial features of Q_n, namely, its f-vector, in particular, the numbers of vertices and facets. It is shown that the latter convergence holds in distribution together with all moments.

Following the setting of [6], we assume that $\Xi_n = \{\xi_1, \dots, \xi_n\}$ is a set of n i.i.d. points distributed in a set $K \subset \mathbb{R}^d$ according to a probability measure μ and define

$$X_n := \bigcap_{i=1}^{n} (K - \xi_i). \tag{1}$$

In difference to [6], it is not assumed that points are uniformly distributed in K and the convexity assumption on K is also dropped. Instead, it is assumed that K is a regular closed set from the extended convex ring, which is the family of locally finite unions of convex bodies, see [5].

The main result states that, after an appropriate multiplicative scaling, the random closed set X_n converges in distribution to the zero cell of a tessellation in $\mathbb{R}^d$, whose distribution is determined by the curvature measure of K and the behaviour of μ near the boundary of K. Some limit results on the expectation of the volume of properly scaled X_n are also presented.

2 Asymptotic Properties of Integrals over Minkowski Differences

By $\mathcal{H}^{d-1}$ we denote the Hausdorff measure of dimension $(d-1)$, and by $\mathcal{H}^d$ or V_d the Lebesgue measure in $\mathbb{R}^d$. We write dx in integrals with respect to the Lebesgue measure.

Let $K \subseteq \mathbb{R}^d$ be a closed set. The metric projection $\xi_K : \mathbb{R}^d \setminus \operatorname{exo}(K) \to K$ is defined by letting $\xi_K(a)$ be the unique nearest point to a from K, where the exoskeleton $\operatorname{exo}(K)$ is the set of points which do not admit a unique nearest point in K. The set $\operatorname{exo}(K)$ is measurable and $V_d(\operatorname{exo}(K)) = 0$, see Sect.2 in [5]. The reduced normal bundle of K is

$$N(K) := \left\{ \left(\xi_K(z), \frac{z - \xi_K(z)}{\|z - \xi_K(z)\|} \right) : z \notin K \cup \operatorname{exo}(K) \right\}.$$

A set $K \subset \mathbb{R}^d$ is called regular closed if it coincides with the closure of its interior. Let $\mathcal{R}$ be the family of regular closed sets. A convex set K is regular closed if and only if its interior is not empty.

A closed set K is said to be gentle if

(G1) $\mathcal{H}^{d-1}(\{a \in B : (a,u) \in N(\partial K), u \in \mathbb{S}^{d-1}\}) < \infty$ for all bounded Borel sets $B \subset \mathbb{R}^d$,

(G2) for $\mathcal{H}^{d-1}$-almost all $a \in \partial K$, there are non-degenerate closed balls B_i and B_o containing a such that $B_i \subset K$ and $\operatorname{Ind} B_o \subset K^c$,

where K^c is the complement of K. A gentle set K is not necessarily regular closed, for example, a singleton is gentle. Each convex body in $\mathbb{R}^d$ is gentle.

Let K be a gentle set. For $\mathcal{H}^{d-1}$-almost all $a \in \partial K$, there exists a unique $u \in \mathbb{S}^{d-1}$ such that $(a,u) \in N(K)$. Let $C_{d-1}(K, \cdot)$ be the image measure of $\mathcal{H}^{d-1}$ on ∂K under the map $a \mapsto (a,u) \in N(K)$, which is defined $\mathcal{H}^{d-1}$-almost everywhere on ∂K and is measurable, see Lemma 6.3 from [3]. The measure $C_{d-1}(K, \cdot)$ is called the curvature measure of K.

We will use the abbreviation for almost all $(a,u) \in N(K)$ in place of for $C_{d-1}(K, \cdot)$-almost all $(a,u) \in N(K)$.

For $L \subseteq \mathbb{R}^d$, its support function is defined as

$$h(L,u) := \sup\{\langle x, u\rangle : x \in L\}, \quad u \in \mathbb{R}^d.$$

For sets K and L in $\mathbb{R}^d$,

$$K \ominus L := \{x \in \mathbb{R}^d : L + x \subseteq K\}$$

is the Minkowski difference of K and L, and the set

$$\check{L} := \{x \in \mathbb{R}^d : -x \in L\}$$

is the reflection of L with respect to the origin.

We impose the following conditions on a finite measure μ supported by K.

(M1) The measure μ has compact support and it is absolutely continuous in a neighbourhood of ∂K with density f.

(M2) There exists an $\alpha > -1$ such that, for almost all $(a, u) \in N(K)$,

$$\lim_{t \downarrow 0} \frac{f(a - tu)}{t^\alpha} = g(a) \in [0, +\infty),$$

where the function g is strictly positive on a subset of ∂K of positive measure and is bounded almost everywhere.

The argument of the limit function g should be the vector $(a, u) \in N(K)$. Since K is gentle, for $\mathcal{H}^{d-1}$-almost all $a \in \partial K$ there is a unique $u \in \mathbb{S}^{d-1}$ such that $(a, u) \in N(K)$, we set $a \in \partial K$ as the argument of g.

As a pointwise limit of measurable functions, g is also measurable. From its definition and since the support of f on K is compact, g has compact support.

Proposition 1. *Let K be a gentle set and let L be a compact set in $\mathbb{R}^d$. If μ is a finite measure on K which satisfies (M1) and (M2), then*

$$\lim_{\varepsilon \downarrow 0} \frac{\mu(K \setminus K \ominus \varepsilon^\gamma L)}{\varepsilon} = \int_{N(K)} g(a) \frac{(h(L, u)^+)^{\alpha+1}}{\alpha + 1} C_{d-1}(K, d(a, u)),$$

where $\gamma = (\alpha + 1)^{-1}$.

Proof. See [10].

3 Limit Theorem for Intersections of Translated Elements of the Extended Convex Ring

A convex body is a non-empty compact convex set. The extended convex ring $\mathcal{U}$ is a family of closed sets which are locally finite unions of convex bodies of $\mathbb{R}^d$, meaning that each compact set intersects at most a finite number of convex bodies that generate an element of $\mathcal{U}$. We assume that the empty set belongs to $\mathcal{U}$. Clearly, the family $\mathcal{U}$ is closed under intersections.

Let K be a non-empty regular closed element of $\mathcal{U}$, and let μ be an absolutely continuous probability measure on K satisfying (M2). Consider the set

$$\Xi_n := \{\xi_1, \dots, \xi_n\}$$

composed of n independent points in K distributed according to μ. Let X_n be a random closed set defined as follows

$$X_n := \bigcap_{i=1}^{n} (K - \xi_i).$$

Let $\mathcal{P}_K := \{(t_i, u_i) : i \geq 1\}$ be a Poisson point process on $(0, \infty) \times \mathbb{S}^{d-1}$ with intensity measure ν, which is the product of an absolutely continuous measure on $(0, \infty)$ with density t^α and a measure $\hat{\nu}$ on $\mathbb{S}^{d-1}$ defined as

$$\hat{\nu}(D) := \int_{\hat{D}(K)} g(a) C_{d-1}(K, d(a, u)), \tag{2}$$

for $D \subset \mathbb{S}^{d-1}$, where $g(a)$ is given by the property (M2) of μ and $\hat{D}(K) := \{(a, u) \in N(K) : u \in D\}$.

The point process $\mathcal{P}_K$ corresponds to the family of hyperplanes $\{x \in \mathbb{R}^d : \langle x, u_i \rangle = t_i\}$, $i \geq 1$, which splits the space into disjoint cells and is said to be a Poisson hyperplane tessellation of $\mathbb{R}^d$, see [9]. The zero cell of this tessellation is the random convex body

$$Z = \bigcap_{i \geq 1} \{x \in \mathbb{R}^d : \langle x, u_i \rangle \leq t_i\}. \tag{3}$$

We now formulate our main result.

Theorem 1. *Assume that μ satisfies (M2) and is absolutely continuous on K. Then*

$$n^\gamma X_n \xrightarrow{d} Z \quad \text{as } n \to \infty,$$

where $\gamma = (\alpha + 1)^{-1}$.

The following result for the proof generalizes a weak convergence criterion for random convex bodies from [7, Proposition 1.8.16] for general regular closed random sets. A random closed set X in $\mathbb{R}^d$ is said to be regular closed if X almost surely belongs to the family $\mathcal{R}$ of regular closed sets, see Sect. 1.1.7 of [7].

Lemma 1. *Let X and X_n, $n \in \mathbb{N}$ be regular closed random sets in $\mathbb{R}^d$. If*

$$\mathbb{P}\{L \subseteq X_n\} \to \mathbb{P}\{L \subseteq X\} \quad \text{as } n \to \infty$$

for all regular closed compact sets L such that $\mathbb{P}\{L \subseteq X\} = \mathbb{P}\{L \subseteq \mathrm{Ind}X\}$, then $X_n \xrightarrow{d} X$ as $n \to \infty$.

Proof. See [10].

Remark 1. By [5, Proposition 2], each regular closed set in the convex ring is gentle. Since properties (G1) and (G2) defining a gentle set are local, each regular closed element of the extended convex ring is also gentle.

The family $\mathcal{R}$ of regular closed sets is not closed under the intersections. The following lemmas show that X_n is a regular closed random set for each $n \in \mathbb{N}$.

Lemma 2. *Let A and B be regular closed elements of $\mathcal{U}$. Then*

$$C := \{x \in \mathbb{R}^d : A \cap (B - x) \notin \mathcal{R}\}$$

is measurable, and its Lebesgue measure is zero.

Proof. See [10].

Lemma 3. *Let μ be an absolutely continuous measure on a regular closed set K. Then X_n is almost surely a regular closed random set for each $n \in \mathbb{N}$.*

Proof. The proof relies on the induction. The step $n = 1$ is clear. Assume that X_{n-1} is a.s. regular closed. Define

$$A_n := \{(x_1, \dots, x_n) \in (\mathbb{R}^d)^n : \cap_{i=1}^n (K - x_i) \in \mathcal{R}\},$$

for $n \in \mathbb{N}$. Then $\mathbb{P}\{(\xi_1, \dots, \xi_{n-1}) \in A_{n-1}\} = 1$. We recall the notation $\Xi_n := \{\xi_1, \dots, \xi_n\}$. Since K and X_{n-1} are almost surely regular closed elements of $\mathcal{U}$, Lemma 2 yields that

$$\begin{aligned}
&\mathbb{P}\{X_n = \mathrm{cl}(\mathrm{Ind}(X_n))\} = \mathbb{E}\Big(\mathbb{P}\{X_n = \mathrm{cl}(\mathrm{Ind}(X_n))|\Xi_{n-1}\}\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_{\Xi_{n-1} \in A_{n-1}} \mathbb{P}\{X_n = \mathrm{cl}(\mathrm{Ind}(X_n))|\Xi_{n-1}\}\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_{\Xi_{n-1} \in A_{n-1}} \mathbb{P}\{\{\xi : X_{n-1} \cap (K - \xi) = \mathrm{cl}(\mathrm{Ind}(X_{n-1} \cap (K - \xi)))\}|\Xi_{n-1}\}\Big) \\
&= 1.
\end{aligned}$$

□

Proof (Proof of Theorom 1). Let L be a compact set in $\mathbb{R}^d$. Then

$$\begin{aligned}
\mathbb{P}\{L \subseteq n^\gamma X_n\} &= \mathbb{P}\left\{n^{-\gamma} L \subseteq (K - \xi_i) \text{ for all } i = 1, \dots, n\right\} \\
&= \left(1 - \mathbb{P}\left\{n^{-\gamma} L \not\subseteq (K - \xi)\right\}\right)^n \\
&= \left(1 - \mathbb{P}\left\{\xi + n^{-\gamma} L \not\subseteq K\right\}\right)^n \\
&= \left(1 - \mathbb{P}\left\{\xi \notin K \ominus n^{-\gamma} L\right\}\right)^n.
\end{aligned}$$

Since K is a regular closed element of $\mathcal{U}$, it is gentle, see Remark 1, and L is compact, Prop. 1 yields that

$$\begin{aligned}
\lim_{n\to\infty} n\mathbb{P}\left\{\xi \notin K \ominus n^{-\gamma} L\right\} &= \lim_{n\to\infty} n\mu(K \setminus K \ominus n^{-\gamma} L) \\
&= \int_{N(K)} g(a) \frac{(h(L,u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a,u)) < \infty.
\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \mathbb{P}\{L \subseteq n^{\gamma} X_n\} = \exp\left(-\int_{N(K)} g(a) \frac{(h(L,u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a,u))\right).$$

Let Z be the zero cell of the tessellation generated by the point process $\mathcal{P}_K$. The random convex set Z satisfies

$$\begin{aligned}\mathbb{P}\{L \subseteq Z\} &= \mathbb{P}\left\{h(L,u)^+ \leq t \text{ for all } (t,u) \in \mathcal{P}_K\right\} \\ &= \exp(-\nu(\{(t,u) \in (0,\infty) \times \mathbb{S}^{d-1} : h(L,u)^+ > t\})) \\ &= \exp\left(-\int_{NK)} g(a) \frac{(h(L,u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a,u))\right).\end{aligned}$$

It follows that,

$$\lim_{n\to\infty} \mathbb{P}\{L \subseteq n^{\gamma} X_n\} = \mathbb{P}\{L \subseteq Z\}.$$

The zero cell Z of the tessellation is a regular closed random set. Indeed, Z is a convex closed subset of $\mathbb{R}^d$. Moreover, all half-spaces from (3) contain the origin in their interior almost surely, so that the interior of Z is not empty with probability one.

The statement follows by Lemma 1, taking into account that $n^{\gamma} X_n$ is regular closed random set by Lemma 3 for each $n \in \mathbb{N}$ and Z is a regular closed random set.

Acknowledgments. The author is grateful to Prof. Ilya Molchanov for all his insightful suggestions and for his patience in correcting the various drafts of this work. This work was supported by the Swiss Enlargement Contribution in the Croatian–Swiss Research Programme framework. (Project number IZHRZ0_180549). The author declares to not have any conflicts of interest. Data sharing is not applicable to this article as no data-sets were generated or analyzed during the current study.

References

1. Fodor, F., Kevei, P., Vígh, V.: On random disc polygons in smooth convex discs. In: Adv. Appl. Probab. 46, 899–918 (2014). https://doi.org/10.1239/aap/1418396236
2. Fodor, F., Papvári, D.I., Vígh, V.: On random approximations by generalized disc-polygons. Mathematika 66, 498–513 (2020) https://doi.org/10.1112/mtk.12027
3. Hug, D., Last, G., Weil, W.: A local Steiner-type formula for general closed sets and applications. Math. Z.. 246, 237–272 (2004). https://doi.org/10.1007/s00209-003-0597-9
4. Kabluchko, Z., Marynych, A., Molchanov, I.: Generalised convexity with respect to families of affine maps. Isr. J. Math, to appear (2023). https://arxiv.org/abs/2202.07887
5. Kiderlen, M., Rataj, J.: On infinitesimal increase of volumes of morphological transforms. Mathematika 53, 103–127 (2007). https://doi.org/10.1112/S002557930000005X

6. Marynych, A., Molchanov, I.: Facial structure of strongly convex sets generated by random samples. Adv. Math. **395**, 108086 (2022). https://doi.org/10.1016/j.aim.2021.108086
7. Molchanov, I.: Theory of Random Sets, 2nd edn. Springer, London (2017)
8. Richey, J., Sarkar, A.: Intersections of random sets. J. Appl. Probab. **59**, 131–151 (2022). https://doi.org/10.1017/jpr.2021.34
9. Schneider, R., Weil, W.: In: Stochastic and Integral Geometry. Springer, Berlin (2008)
10. Visonà, T.: Intersections of randomly translated sets (2023). https://arxiv.org/abs/2308.04242

Hierarchical Variable Clustering Based on Measures of Predictability

Yuping Wang(✉) and Sebastian Fuchs

Department for Artificial Intelligence and Human Interfaces, University of Salzburg, Salzburg, Austria

yuping.wang@plus.ac.at, sebastian.fuchs@plus.ac.at

Abstract. Within a portfolio of random variables, complex structures and associations can occur that influence the overall behaviour of the portfolio. For revealing and identifying such interrelationships, a variable cluster analysis can be performed as a pre-processing step in a data analysis. Motivated by recent developments in the field of hierarchical variable clustering, we here present a short and concise summary of the article *Hierarchical variable clustering based on the predictive strength between random vectors* [5], in which a variable clustering procedure is introduced that is based on a measure of predictability.

Keywords: Agglomerative hierarchical variable clustering · Dissimilarity · Directed dependence · (mutual) perfect dependence

1 Introduction

A cluster analysis for a portfolio $\mathcal{X} = \{X_1, \ldots, X_m\}$ of $m \geq 3$ random variables aims at grouping objects that exhibit a similar behaviour. Thereby, similarity has long been treated as a characteristic being restricted to pairs of variables. A similarity degree for two groups, each possibly consisting of more than one variable, has then been determined using so-called linkage functions, with the pairwise similarity degrees concerning the variables in the different groups being merged in an appropriate fashion. However, this approach has the disadvantage that only pairwise information among the variables is taken into account and higher-dimensional dependence is completely ignored.

To remedy this drawback, the authors in [4,6] propose dependence concepts that allow for quantifying multivariate similarity. This involves multivariate similarity based on 'concordance' and 'tail-dependence' in [6] and on Φ-dependence in [4]. In contrast, the authors in [5] treat 'similarity' as equivalent to the concept of 'predictability', thus assigning a high similarity degree to two disjoint groups of variables if the variables within one of the groups contain high predictive information about the variables within the other group and/or vice versa; cf. [5]. The following two concepts are used to characterize full predictability in either one or both directions:

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 548–553, 2024.
https://doi.org/10.1007/978-3-031-65993-5_67

- A set $\mathbb{Y} = \{Y_1, \ldots, Y_q\}$ of $q \geq 1$ random variables is said to be *perfectly dependent* on the set of random variables $\mathbb{X} = \{X_1, \ldots, X_p\}$, $p \geq 1$, if, for every $k \in \{1, \ldots, q\}$, there exists some measurable function f_k such that $Y_k = f_k(X_1, \ldots, X_p)$ almost surely.

- The two sets $\mathbb{X} = \{X_1, \ldots, X_p\}$, $p \geq 1$, and $\mathbb{Y} = \{Y_1, \ldots, Y_q\}$, $q \geq 1$, are said to be *mutually perfectly dependent* if $\mathbb{Y}$ is perfectly dependent on $\mathbb{X}$ and $\mathbb{X}$ is perfectly dependent on $\mathbb{Y}$.

Clearly, mutual perfect dependence implies perfect dependence. It is worth mentioning that perfect dependence of $\mathbb{Y}$ on $\mathbb{X}$ is an order-free concept of directed dependence in the sense that the variables within $\mathbb{X}$ provide full information about the variables within $\mathbb{Y}$, regardless of the order of the variables within $\mathbb{X}$ and $\mathbb{Y}$.

Throughout, we assume that the random variables are all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and are non-degenerate. Any subset of $\mathcal{X}$ will be denoted by upper-case black-board letters, e.g., $\mathbb{X}$, and let $\mathcal{P}_0(\mathcal{X})$ denote the set of all non–empty subsets of $\mathcal{X}$. Given a subset $\mathbb{X} = \{X_1, \ldots, X_p\} \in \mathcal{P}_0(\mathcal{X})$ with $p \leq m$, we indicate by $\vec{\mathbb{X}}$ a vector representation of $\mathbb{X}$. Clearly, the vector representation of any $\mathbb{X}$ need not be unique.

For quantifying the degree of 'predictability', the authors in [5] propose to use the so-called *simple extension of Azadkia & Chatterjee's rank correlation* T^q introduced in [1]. For the q-dimensional random vector $\vec{\mathbb{Y}}$, $q \geq 1$, given the p-dimensional random vector $\vec{\mathbb{X}}$, T^q is defined by

$$T^q(\vec{\mathbb{Y}}|\vec{\mathbb{X}}) := 1 - \frac{q - \sum_{i=1}^{q} T\big(Y_i|(\vec{\mathbb{X}}, Y_{i-1}, \ldots, Y_1)\big)}{q - \sum_{i=1}^{q} T\big(Y_i|(Y_{i-1}, \ldots, Y_1)\big)}$$

with $T(Y_1|\emptyset) := 0$ and T denoting the *simple measure of conditional dependence* introduced in [3] and given by

$$T(Y|\vec{\mathbb{X}}) := \frac{\int_{\mathbb{R}} \mathrm{Var}\big(\mathbb{P}(Y \geq y|\vec{\mathbb{X}})\big)\, \mathrm{d}\mathbb{P}^Y(y)}{\int_{\mathbb{R}} \mathrm{Var}(\mathbb{1}_{\{Y \geq y\}})\, \mathrm{d}\mathbb{P}^Y(y)}\,.$$

At first glance, T as defined above quantifies the variability of the conditional distributions $\mathbb{P}^{Y|\vec{\mathcal{X}}}$, but can also be interpreted as a coefficient indicating how sensitive the conditional distribution $\mathbb{P}^{Y|\vec{\mathcal{X}}=\mathbf{x}}$ of Y given $\vec{\mathbb{X}} = \boldsymbol{x}$ is on $\boldsymbol{x}$; cf. [2].

As shown in [1], $T^q(\vec{\mathbb{Y}}|\vec{\mathbb{X}})$ is invariant with respect to permutations of the variables within $\mathbb{X} = \{X_1, \ldots, X_p\}$, and permutation invariance with respect to the variables within $\mathbb{Y} = \{Y_1, \ldots, Y_q\}$ can be achieved by defining the map

$$\kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}}) := \frac{1}{q!} \sum_{\sigma \in S_q} T^q(\sigma(\vec{\mathbb{Y}})|\vec{\mathbb{X}}) \tag{1}$$

where $\sigma(\vec{\mathbb{Y}}) := (Y_{\sigma_1}, \ldots, Y_{\sigma_q})$ for all $\sigma = (\sigma_1, \ldots, \sigma_q) \in S_q$. $\kappa^{q|p}$ is a *measure of predictability* for the q-dimensional random vector $\vec{\mathbb{Y}}$ given the p-dimensional random vector $\vec{\mathbb{X}}$, i.e., it fulfills the following properties; see, e.g., [5]:

- $0 \leq \kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}}) \leq 1$.
- $\kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}}) = 0$ iff the variables in $\mathbb{Y}$ are independent of those in $\mathbb{X}$.
- $\kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}}) = 1$ iff $\mathbb{Y}$ is perfectly dependent on $\mathbb{X}$.

2 Hierarchical Variable Clustering Based on (Mutual) Perfect Dependence

An agglomerative hierarchical variable clustering procedure based on (mutual) perfect dependence requires the notion of a so-called dissimilarity function: A (p, q)-*dissimilarity function* $d^{p,q} : L^0(\mathbb{R}^p) \times L^0(\mathbb{R}^q) \mapsto [0, \infty)$ is a mapping that assigns a non-negative value to every pair of a p-dimensional and a q-dimensional random vector, where $L^0(\mathbb{R}^d)$ denotes the space of all d-dimensional random vectors, thereby fulfills the subsequent properties:

(d1) Symmetry and permutation invariance: The identities

$$d^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = d^{q,p}(\vec{\mathbb{Y}}, \vec{\mathbb{X}}) \qquad d^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = d^{p,q}(\sigma_p(\vec{\mathbb{X}}), \sigma_q(\vec{\mathbb{Y}}))$$

hold for all $\sigma_p \in S_p$ and $\sigma_q \in S_q$.

(d2) Law-invariance: $d^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = d^{p,q}(\vec{\mathbb{X}}_1, \vec{\mathbb{Y}}_1)$ for all $(\vec{\mathbb{X}}, \vec{\mathbb{Y}})(\vec{\mathbb{X}}_1, \vec{\mathbb{Y}}_1)$ such that $(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) \stackrel{d}{=} (\vec{\mathbb{X}}_1, \vec{\mathbb{Y}}_1)$.

(d3) Multivariate similarity

(A) based on perfect dependence:
$d^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 0$ iff either $\mathbb{Y}$ is perfectly dependent on $\mathbb{X}$ *or* $\mathbb{X}$ is perfectly dependent on $\mathbb{Y}$.

(B) based on mutual perfect dependence:
$d^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 0$ iff both $\mathbb{Y}$ is perfectly dependent on $\mathbb{X}$ *and* $\mathbb{X}$ is perfectly dependent on $\mathbb{Y}$.

We refer to [5,6] for a detailed discussion of these properties.

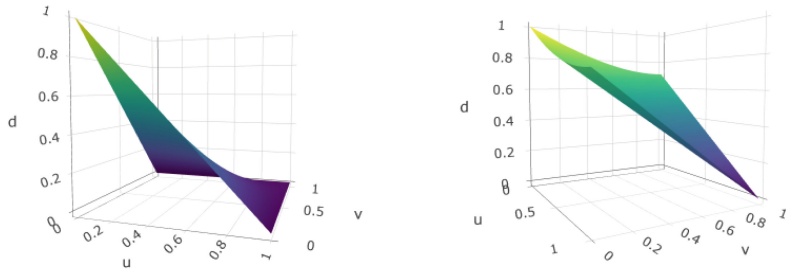
For constructing dissimilarity functions based on (mutual) perfect dependence, the individual degrees of predictability $\kappa^{p|q}(\vec{\mathbb{X}}|\vec{\mathbb{Y}})$ and $\kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}})$ need to be aggregated appropriately. Therefore, consider the mappings $d_A^{p,q}, d_B^{p,q} : L^0(\mathbb{R}^p) \times L^0(\mathbb{R}^q) \mapsto [0, \infty)$ given by

(A) Perfect dependence:

$$d_A^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) := \left(1 - \kappa^{p|q}(\vec{\mathbb{X}}|\vec{\mathbb{Y}})\right)\left(1 - \kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}})\right), \tag{2}$$

(B) Mutual perfect dependence:

$$d_B^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) := \sqrt{\frac{\left(1 - \kappa^{p|q}(\vec{\mathbb{X}}|\vec{\mathbb{Y}})\right)^2 + \left(1 - \kappa^{q|p}(\vec{\mathbb{Y}}|\vec{\mathbb{X}})\right)^2}{2}}. \tag{3}$$



(a) Type A (b) Type B

Fig. 1. Graphs of the dissimilarity functions $d_A^{p,q}$ of type (A) (left panel) and $d_B^{p,q}$ of type (B) (right panel).

Then $d_A^{p,q}$ and $d_B^{p,q}$ are proper dissimilarity function based on perfect dependence and mutual perfect dependence, respectively; see Fig. 1 for an illustration. For a detailed discussion of dissimilarity functions of type (A) and (B), alternative choices and additional properties, we refer to [5].

Theorem 1.

(A) The dissimilarity function $d_A^{p,q}$ of type (A) defined in (2) fulfills

1. $d_A^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) \in [0, 1]$.
2. $d_A^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 0$ *iff either* $\mathbb{Y}$ *is perfectly dependent on* $\mathbb{X}$ or $\mathbb{X}$ *is perfectly dependent on* $\mathbb{Y}$.
3. $d_A^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 1$ *iff the variables in* $\mathbb{Y}$ *are independent of those in* $\mathbb{X}$.

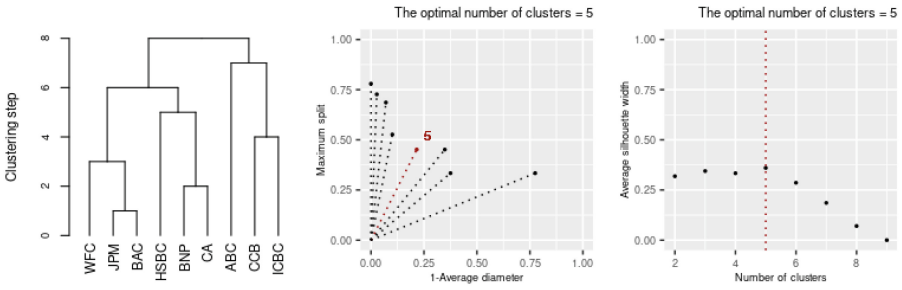
(B) The dissimilarity function $d_B^{p,q}$ of type (B) defined in (3) fulfills

1. $d_B^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) \in [0, 1]$.
2. $d_B^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 0$ *iff* $\mathbb{Y}$ *is perfectly dependent on* $\mathbb{X}$ and $\mathbb{X}$ *is perfectly dependent on* $\mathbb{Y}$.
3. $d_B^{p,q}(\vec{\mathbb{X}}, \vec{\mathbb{Y}}) = 1$ *iff the variables in* $\mathbb{Y}$ *are independent of those in* $\mathbb{X}$.

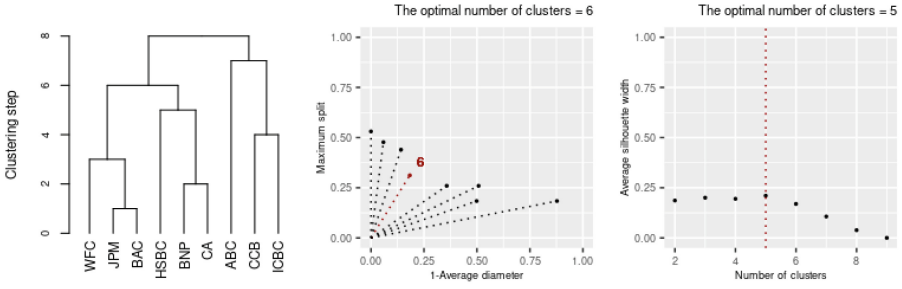
3 Data Example

For illustrating the excellent performance of the here proposed hierarchical clustering procedure, we performed a cluster analysis investigating the interconnectedness of the three largest banks in each of the U.S. (US), Europe (EU), and Asia and Pacific (AP). For the U.S. this includes Wells Fargo (WFC), Bank of America (BAC) and JP Morgan (JPM), while the three largest banks in Europe comprise HSBC in the U.K., BNP Paribas (BNP) and Crédit Agricole (CA). In Asia and the Pacific, the three largest banks are the Agricultural Bank of China (ABC), China Construction Bank Corporation (CCB), and Industrial and Commercial Bank of China (ICBC).

For both, the hierarchical clustering procedure based on perfect dependence (type (A)) and mutual perfect dependence (type (B)). We employed a sample of



(a) Type (A)



(b) Type (B)

Fig. 2. Dendrograms (left) and the corresponding trade-off between average diameter and maximum split (middle) & Silhouette coefficient (right) for the hierarchical clustering of 9 banks using type (A) and (B) dissimilarity functions.

daily log-returns of the nine banks' stock data (in US dollar) from April 7, 2011 to December 14, 2022, thereby assuming that the log-returns are i.i.d., which is a standard assumption for stock data; see, e.g., [9].

Figure 2 depicts the dendrograms obtained by employing type (A) and (B) dissimilarity functions, together with the optimal number of clusters determined via two different optimality criteria (1) a trade-off between average diameter measuring intra-cluster similarity and maximum split measuring inter-cluster similarity and (2) the silhouette coefficient; cf. [5].

We find that the two dendrograms are identical and that the banks are clustered according to the three regions U.S., Europe and Asian and Pacific. Following both optimality criteria, the estimated optimal number of clusters for type (A) equals 5 with the optimal partition: cluster 1 = {WFC, JPM, BAC}, cluster 2 = {HSBC}, cluster 3 = {BNP, CA}, cluster 4 = {CCB, ICBC}, and cluster 5 = {ABC}. Interestingly, the U.K. bank HSBC is not clustered together with the mainland European banks BNP and CA. The estimated optimal number of clusters for type (B) ranges between 5 and 6 with a similar optimal partition.

Acknowledgments. Both authors gratefully acknowledge the support of the Austrian Science Fund (FWF) project P 36155-N *ReDim: Quantifying Dependence via*

Dimension Reduction and the support of the WISS 2025 project 'IDA-lab Salzburg' (20204-WISS/225/197–2019 and 20102-F1901166-KZP).

References

1. Ansari, J., Fuchs, S.: A simple extension of Azadkia & Chatterjee's rank correlation to a vector of endogenous variables (2023). Available at arxiv.org/abs/2212.01621
2. Ansari, J., Langthaler, P.B., Fuchs, S., Trutschnig, W.: Quantifying and estimating dependence via sensitivity of conditional distributions (2023). Available at arxiv.org/abs/2308.06168
3. Azadkia, M., Chatterjee, S.: A simple measure of conditional dependence. Ann. Statist. **49**(6), 3070–3102 (2021). https://doi.org/10.1214/21-AOS2073
4. de Keyser, S., Gijbels, I.: Hierarchical variable clustering via copula-based divergence measures between random vectors, Int. J. Approx. Reason. **165**, 109090 (2024). https://doi.org/10.1016/j.ijar.2023.109090
5. Fuchs, S., Wang, Y.: Hierarchical variable clustering based on the predictive strength between random vectors (2023). Available at arxiv.org/abs/2312.16544
6. Fuchs, S., Di Lascio, F.M.L., Durante, F.: Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Comput. Statist. Data Anal. **159**, 107201 (2021). https://doi.org/10.1016/j.csda.2021.107201
7. Kaufman, L.: Finding Groups in Data. Wiley & Sons (1990)
8. Kojadinovic, I.: Agglomerative hierarchical clustering of continuous variables based on mutual information. Comput. Statist. Data Anal. **46**(2), 269–294 (2004). https://doi.org/10.1016/S0167-9473(03)00153-1
9. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, NJ (2015)

Large Deviation Theory Without Probability Measures

José M. Zapata(✉)

Dpto. de Estadística e Investigación Operativa, Universidad de Murcia, 30100 Espinardo, Murcia, Spain

jmzg1@um.es

Abstract. Large deviation theory studies the asymptotic tail behaviour of sequences of probability distributions. A series of recent results have generalized the basic principles and theorems of this theory to maxitive translation invariant previsions, stripping away the probabilistic aspects of this theory. In this note we survey these results, and explain their connection to large deviation theory.

Keywords: Large deviation theory · Maxitive translation invariant previsions · Convex integral · Large deviation principle · Laplace principle

1 Introduction

Large deviation theory, a branch of probability theory, covers the asymptotic behavior of sequences of probability distributions. Its foundations stem from early developments in ruin theory by Cramér [4,5], while Varadhan and Donsker [7,17] formalized the modern framework. For further historical context and details, we refer to the monograph [6].

Consider a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables taking values in a topological space E. Large deviation theory aims to establish the exponential decay of the probability $\mathbb{P}(X_n \in A)$ for certain Borel sets A by finding bounds for the lower and upper large deviation estimators:

$$\underline{J}_A := \liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}(X_n \in A) \quad \text{and} \quad \overline{J}_A := \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}(X_n \in A).$$

In particular, the sequence $(X_n)_{n\in\mathbb{N}}$ is said to satisfy the *large deviation principle* (LDP) with rate function $I\colon E \to [0,\infty]$ if

$$-\inf_{x\in\mathrm{int}(A)} I(x) \le \underline{J}_A \le \overline{J}_A \le -\inf_{x\in\mathrm{cl}(A)} I(x) \tag{1}$$

for all Borel sets $A \subset E$.[1] Varadhan's integral lemma asserts that a sequence $(X_n)_{n\in\mathbb{N}}$ that satisfies the LDP with rate function $I(\cdot)$ also satisfies the *Laplace*

[1] We denote by $\mathrm{int}(A)$ and $\mathrm{cl}(A)$ the topological interior and closure of $A \subset E$, respectively.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 554–561, 2024.
https://doi.org/10.1007/978-3-031-65993-5_68

principle (LP) with rate function $I(\cdot)$, that is,

$$\psi(f) = \sup_{x \in E} \{f(x) - I(x)\} \tag{2}$$

for all $f \in C_b(E)$, where $\psi(f) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{nf(X_n)}]$ is the *Varadhan functional*.[2] Other basic results such as Bryc's converse of the Varadhan lemma, and Gärtner-Ellis theorem provide sufficient conditions for both the LDP and LP.

Several works have explored large deviation theory based on other classes of functionals rather than the Varadhan functional, see e.g., [1,3,8,10,13,16,19]. Motivated by these works, Kupper and Zapata [11] (see also [12,20]) have recently developed a structural theory of translation invariant maxitive previsions, showing that the properties of these functions are enough to establish generalized versions of Varadhan's integral lemma and its Bryc's converse, and the Gärtner-Ellis theorem. The approach in [11,20] fully relies on topological and order properties and, in particular, an underlying probability space is not needed. More specifically, [11,20] exploit the fact that every maxitive translation invariant prevision admits a representation as a convex integral (see [2]), in combination with duality bounds for the convex integral as those established in [12].

In this note, we provide a survey of the main results and ideas in [11,12,20] in a simple and unified setting. Section 2 covers a review of the convex integral, its properties and duality bounds. In Sect. 3, we revisit the main results from [11,20] focusing on versions of Varadhan's lemma and its Bryc's converse, and Gärtner-Ellis theorem.

2 Convex Integral: Duality Bounds and Representation

Throughout this note, let E be a regular Hausdorff topological space with Borel σ-algebra $\mathcal{B}(E)$. A set function $J\colon \mathcal{B}(E) \to [-\infty, 0]$ is said to be a *concentration* if $J_\emptyset = -\infty$, $J_E = 0$, and $J_A \leq J_B$ whenever $A \subset B$. We say that J is *maxitive* if $J_{A\cup B} \leq J_A \vee J_B$ for all $A, B \in \mathcal{B}(E)$. Denote by $B(E)$ the set of all Borel measurable functions $f\colon E \to \mathbb{R} \cup \{-\infty\}$. The *convex integral* of $f \in B(E)$ with respect to the concentration J is defined as

$$\phi_J(f) = \sup_{c\in\mathbb{R}} \left\{c + J_{\{f>c\}}\right\}. \tag{3}$$

Remark 1. The convex integral (3) was introduced in [2] and can be obtained as a transformation of the Shilkret integral [15]. We point out that Puhalskii [14] described the LPD and LP in terms of the Shilkret integral.

The convex integral satisfies the following properties, see e.g., [2,12,20]:

[2] Here, $C_b(E)$ denotes the set of all bounded continuous functions from E to $\mathbb{R}$. The limit in $\psi(f)$ exists for all $f \in C_b(E)$ if $(X_n)_{n\in\mathbb{N}}$ satisfies the LDP.

1. $\phi_J(\chi_A) = J_A$ for all $A \in \mathcal{B}(E)$, where χ_A denotes the function that takes the value 0 on A and the value $-\infty$ on A^c.
2. $\phi_J(0) = 0$,
3. $\phi_J(f + c) = \phi_J(f) + c$ for all constant $c \in \mathbb{R}$,
4. $\phi_J(f) \leq \phi_J(g)$ whenever $f \leq g$.

2.1 Duality Bounds

In this subsection, let $J\colon \mathcal{B}(E) \to [-\infty, 0]$ be a concentration. Denote by $L(E)$ the set of all lower semicontinuous functions $f\colon E \to \mathbb{R} \cup \{-\infty\}$, and by $U(E)$ the set of all upper semicontinuous functions $f\colon E \to \mathbb{R} \cup \{-\infty\}$.

The results that we state next were provided in [12] in a more general setting. An exposition in the present setting can be found in [20, Sect. 2].

Theorem 1. *Let $I\colon E \to [0, \infty]$ be a function. Then, the following equivalences hold. First,*

$$-\inf_{x \in O} I(x) \leq J_O \quad \text{for all open sets } O \subset E \tag{4}$$

if and only if

$$\phi_J(f) \geq \sup_{x \in E}\{f(x) - I(x)\} \quad \text{for all } f \in L(E). \tag{5}$$

Second,

$$J_C \leq -\inf_{x \in C} I(x) \quad \text{for all closed sets } C \subset E \tag{6}$$

if and only if

$$\phi_J(f) \leq \sup_{x \in E}\{f(x) - I(x)\} \quad \text{for all } f \in U(E). \tag{7}$$

The minimal rate function $I_{\min}\colon E \to [0, \infty]$ associated with the concentration J is defined as

$$I_{\min}(x) = \sup_{f \in L(E)} \{f(x) - \phi_J(f)\}. \tag{8}$$

Proposition 1. *$I_{\min}$ is the smallest mapping $I\colon E \to [0, \infty]$ which satisfies the equivalent inequalities (4) and (5).*

The minimal rate function has the following representation.

Proposition 2. *For all $x \in E$, it holds*

$$-I_{\min}(x) = \inf_{U \in \mathcal{U}_x} J_U,$$

where $\mathcal{U}_x$ is any base of open neighborhoods of $x \in E$.

Definition 1. *We say that J is* tight *if for every $n \in \mathbb{N}$ there exists a compact set $K \subset E$ such that $J_{K^c} < -n$.*

Proposition 3. *If J is tight and maxitive, then $I_{\min}$ satisfies the inequalities* (4), (5), (6), *and* (7).

Remark 2. The assumption of maxitivity in Proposition 3 can be relaxed to a weaker topological form of maxitivity, as stated in [12, Corollary 3.10] and [20, Proposition 2.6].

For a proof of the following result, we refer to [12, Lemma 5.4].[3]

Proposition 4. *If E is completely regular, then*

$$I_{\min}(x) = \sup_{f \in C_b(E)} \{f(x) - \phi_J(f)\} \quad \text{for all } x \in E.$$

2.2 Integral Representation of Maxitive Previsions

A *translation invariant prevision* is a function $\phi\colon B(E) \to [-\infty, \infty]$ satisfying:

(N) Normalization: $\phi(0) = 0$,
(M) Monotonicity: $\phi(f) \leq \phi(g)$ whenever $f \leq g$,
(T) Translation invariance: $\phi(f + c) = \phi(f) + c$ for all $c \in \mathbb{R}$.

Among the main examples of translation invariant previsions we find coherent previsions in imprecise probability [18], and monetary risk measures in risk analysis [9]. We say that a prevision ϕ is *maxitive* if it satisfies

$$\phi(f \vee g) \leq \phi(f) \vee \phi(g) \quad \text{for all } f, g \in B(E).$$

As seen in the previous section, the convex integral (3) is a translation invariant prevision which is maxitive if the corresponding concentration is maxitive. Next, we cover the converse direction, examining the conditions under which a translation invariant prevision can be expressed as a convex integral. Denote by $\bar{B}(E)$ the set of all $f \in B(E)$ such that f is bounded from above. The following result was obtained in [2, Corollary 6] (see also [20, Theorem 3.1]).

Theorem 2. *Let $\phi\colon B(E) \to [-\infty, \infty]$ be a maxitive translation invariant prevision, and let $J_A = \phi(\chi_A)$ for all $A \in \mathcal{B}(A)$. Then, J is a maxitive concentration, and*

$$\phi(f) = \phi_J(f)$$

for all $f \in \bar{B}(E)$.

Theorem 2 is extended in [20, Theorem 3.5] to unbounded functions as follows. Given a prevision $\phi\colon B(E) \to [-\infty, \infty]$, we define the set

$$B_\phi(E) := \{f \in B(E)\colon \text{ there exists } t > 1 \text{ such that } \phi(tf) < \infty\}.$$

[3] Lemma 5.4 in [12] is proved for general preordered topological spaces. We point out that [12, Lemma 5.4] is proven under Assumption 5.1 in [11]. As explained in [12, Remark 5.2], Assumption 5.1 corresponds to complete regularity in the case of the trivial preorder.

Theorem 3. *Let $\phi\colon B(E) \to [-\infty,\infty]$ be a translation invariant prevision, and let $J_A = \phi(\chi_A)$ for all $A \in \mathcal{B}(E)$. If ϕ is maxitive, then $\phi(f) = \phi_J(f)$ for all $f \in B_\phi(E)$.*

Remark 3. By relaxing the maxitivity assumption to a weak topological form of maxitivity, it is still possible to represent a prevision in terms of the convex integral as in Theorems 2 and 3 on the continuous functions in $B(E)$, see [12, Theorem 3.3] and [20, Theorem 3.5].

Example 1. Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of E-valued random variables defined on a probability space $(\Omega,\Sigma,\mathbb{P})$.

1. The upper Varadhan functional

$$\overline{\psi}(f) := \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\left[e^{nf(X_n)}\right]$$

is maxitive and $\overline{\psi}(\chi_A) = \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(X_n \in A)$ is the upper large deviation estimator, see [11,12,20]. As a consequence of Theorem 3,

$$\overline{\psi}(f) = \phi_{\overline{J}}(f) \quad \text{for all } f \in B_{\overline{\psi}}(E),$$

where $\overline{J}_A := \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(X_n \in A)$.

2. Consider a set $\mathcal{P}$ of probability measures on Σ. Then, the upper robust Varadhan functional

$$\overline{\psi}_{\mathcal{P}}(f) := \limsup_{n\to\infty} \tfrac{1}{n} \log \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[e^{nf(X_n)}\right]$$

is maxitive and $\overline{\psi}_{\mathcal{P}}(\chi_A) = \limsup_{n\to\infty} \frac{1}{n}\log \sup_{P\in\mathcal{P}} P(X_n \in A)$, see [11,12, 20]. As a consequence of Theorem 3,

$$\overline{\psi}_{\mathcal{P}}(f) = \phi_{\overline{J}}(f) \quad \text{for all } f \in B_{\overline{\psi}_{\mathcal{P}}}(E),$$

where $\overline{J}_A := \limsup_{n\to\infty} \frac{1}{n}\log \sup_{P\in\mathcal{P}} P(X_n \in A)$.

3 Large Deviation Theory Without Probability Measures

In this section, we survey some generalizations of basic results of large deviation theory to pairs of previsions obtained in [11,20]. Throughout this section, we consider two translation invariant previsions $\underline{\phi}, \overline{\phi}\colon B(E) \to [-\infty,\infty]$ such that $\overline{\phi}$ is maxitive, and $\underline{\phi}(f) \le \overline{\phi}(f)$ for all $f \in B(E)$. We consider the concentrations $\underline{J}, \overline{J}\colon \mathcal{B}(E) \to [-\infty, 0]$ given by $\underline{J}_A = \underline{\phi}(\chi_A)$ and $\overline{J}_A = \overline{\phi}(\chi_A)$, and the spaces of continuous functions $C(E) = \underline{L}(E) \cap \overline{U}(E)$ and $C_{\overline{\phi}}(E) = B_{\overline{\phi}}(E) \cap C(E)$.

Definition 2. *Suppose that $I\colon E \to [0,\infty]$ is a rate function.*

– *We say that the pair* $\underline{\phi}, \overline{\phi}$ *satisfies the* large deviation principle *(LDP) with rate function* $I(\cdot)$ *if*

$$-\inf_{x\in\mathrm{int}(A)} I(x) \le \underline{J}_A \le \overline{J}_A \le -\inf_{x\in\mathrm{cl}(A)} I(x) \quad \textit{for all } A \in \mathcal{B}(E).$$

– *We say that the pair* $\underline{\phi}, \overline{\phi}$ *satisfies the* Laplace principle *(LP) with rate function* $I(\cdot)$ *if*

$$\underline{\phi}(f) = \overline{\phi}(f) = \sup_{x\in E}\{f(x) - I(x)\} \quad \textit{for all } f \in C_{\overline{\phi}}(E).$$

Remark 4. The LDP and LP for pairs of translation invariant previsions on $C_b(E)$ are introduced in [11, Sect. 5]. In [11], the LP is defined on $C_b(E)$. Here, as in [20, Definition 4.3], we consider the larger space $C_{\overline{\phi}}(E)$.

3.1 Varadhan's Integral Lemma

The following result was obtained in [11, Proposition 5.2]. Here, we state a slightly more general version, which is obtained by combining the duality bounds in Theorems 1 and 2. The detailed proof can be found in [20, Proposition A.1].

Theorem 4. *Let* $I\colon E \to [0,\infty]$ *be a rate function. If the pair* $\underline{\phi}, \overline{\phi}$ *satisfies the LDP with rate function* $I(\cdot)$*, then the pair* $\underline{\phi}, \overline{\phi}$ *satisfies the LP with rate function* $I(\cdot)$*.*

Remark 5. The upper Varadhan functional $\overline{\psi}$ is maxitive, see Example 1. Thus, Theorem 4 amounts to Varadhan's integral lemma (see [6, Theorem 4.3.1]) in this particular case of the lower/upper Varadhan's functionals $\underline{\psi}, \overline{\psi}$.[4]

3.2 Bryc's Converse Varadhan Lemma

The following result was provided in [11, Theorem 4.1]. Here, we state a slightly different version for which we provide a short proof.

Theorem 5. *Suppose that* E *is completely regular,* $\overline{J}$ *is tight, and*

$$\overline{\phi}(f) = \underline{\phi}(f) \quad \textit{for all } f \in C_b(E). \tag{9}$$

Then, the pair $\underline{\phi}, \overline{\phi}$ *satisfies both the LDP and LP with rate function*

$$I(x) = \sup_{f\in C_b(E)} \{f(x) - \overline{\phi}(f)\}.$$

Proof. We denote by $\underline{I}, \overline{I}$ the respective minimal rate functions of $\underline{\phi}, \overline{\phi}$ defined as in (8). Since E is completely regular, as a consequence of Proposition 4, Theorem 2, we have that

$$I(x) = \overline{I}(x) = \sup_{f\in C_b(E)} \{f(x) - \overline{\phi}(f)\} = \sup_{f\in C_b(E)} \{f(x) - \underline{\phi}(f)\} = \underline{I}(x),$$

where we have applied (9). Since $\overline{J}$ is tight, the conclusion follows from Proposition 3.

[4] The lower Varadhan's functional $\underline{\psi}$ is defined similarly as $\overline{\psi}$ replacing the limit superior by the limit inferior.

3.3 Gärtner-Ellis Theorem

In this subsection, we make the following assumption:

Assumption 1. $\underline{\phi}(f \vee g) \leq \underline{\phi}(f) \vee \overline{\phi}(g)$ for all $f, g \in B(E)$.

Let $\mathcal{F}$ be any given nonempty set of continuous real-valued functions on E. We define the *conjugate* $\overline{\phi}^*_{\mathcal{F}} \colon E \to [-\infty, \infty]$ of $\overline{\phi}$ with respect to $\mathcal{F}$ as

$$\overline{\phi}^*_{\mathcal{F}}(x) := \sup_{f \in \mathcal{F}} \{f(x) - \overline{\phi}(f)\}.$$

Definition 3. *We say that $x \in E$ is an* $\mathcal{F}$-exposed point *of $\overline{\phi}^*_{\mathcal{F}}$ if there exists $f \in \mathcal{F}$ such that*

$$f(y) - \overline{\phi}^*_{\mathcal{F}}(y) < f(x) - \overline{\phi}^*_{\mathcal{F}}(x) \quad \text{for all } y \neq x.$$

In that case, we say that f is an exposing function *for x. We denote by $\mathscr{E}$ the set of all $\mathcal{F}$-exposed points $x \in E$ of $\overline{\phi}^*_{\mathcal{F}}$ which admit an exposing function $f \in \mathcal{F}$ such that*

$$\overline{\phi}(f) = \underline{\phi}(f), \quad \text{and} \quad f \in B_{\overline{\phi}}(E). \tag{10}$$

The following result was provided in [20, Theorem 4.7].

Theorem 6. *Suppose that $\overline{J}$ is tight. Then:*

1. *For every closed set $C \subset E$, we have the upper bound*

$$\overline{J}_C \leq -\inf_{y \in C} \overline{\phi}^*_{\mathcal{F}}(x).$$

2. *For every open set $O \subset E$, we have the lower bound*

$$-\inf_{y \in O \cap \mathscr{E}} \overline{\phi}^*_{\mathcal{F}}(x) \leq \underline{J}_O.$$

3. *If, moreover,*

$$\inf_{x \in O} \overline{\phi}^*_{\mathcal{F}}(x) = \inf_{x \in O \cap \mathscr{E}} \overline{\phi}^*_{\mathcal{F}}(x) \quad \text{for every } O \subset E \text{ open,} \tag{11}$$

then the pair $\underline{\phi}, \overline{\phi}$ verifies the LDP and LP with rate function $\overline{\phi}^*_{\mathcal{F}}$.

Remark 6. The upper/lower Varadhan functionals $\overline{\psi}, \underline{\psi}$ verify Assumption 1 above, see [20]. Consequently, Theorem 6 is applicable in this particular case. It seems that before [20], this result was only known in the case that $\mathcal{F}$ is the dual topological space of a locally convex topological vector space. Indeed, this result in the realm of large deviation theory is known as the Gärtner-Ellis theorem, see e.g., [6, Theorem 4.5.20]. In contrast to the Gärtner-Ellis theorem, Theorem 6 allows for arbitrary choices of the set $\mathcal{F}$, covering situations where the Gärtner-Ellis theorem fails. For instance, [20, Example 5.5] shows a situation where the LDP and the LP follow from Theorem 6 but are not covered by the Gärtner-Ellis theorem.

References

1. Bell, H., Bryc, W.: Variational representations of Varadhan functionals. Proc. Amer. Math. Soc. **129**, 2119–2125 (2001)
2. Cattaneo, M.: On maxitive integration. Fuzzy Sets Syst. **304**, 65–81 (2016)
3. Chen, Z., Feng, X.: Large deviation for negatively dependent random variables under sublinear expectation. Commun. Stat. Theory Methods **45**(2), 400–412 (2015)
4. Cramér, H.: Sur un nouveau théoreme-limite de la théorie des probabilités. Actual. Sci. Ind. **736**, 5–23 (1938)
5. Cramér, H.: Historical review of Filip Lundberg's works on risk theory. Scand. Actuar. J. 6–12 (1969)
6. Dembo, A., Zeitouini, O.: In: Large Deviations Techniques and Applications. Springer, Berlin (2010)
7. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time I. Commun. Pure Appl. Math. **28**, 1–47 (1975)
8. Föllmer, H., Knispel, T.: Entropic risk measures: Coherence versus convexity, model ambiguity, and robust large deviations. Stoch. Dyn. **11**, 333–351 (2011)
9. Föllmer, H., Schied, A.: In: Stochastic Finance. An Introduction in Discrete Time, extended edn. Walter de Gruyter & Co., Berlin (2011)
10. Hu, F.: On Cramér's theorem for capacities. C. R. Math. **348**(17–18), 1009–1013 (2010)
11. Kupper, M., Zapata, J.M.: Large deviations built on max-stability. Bernoulli **27**(2), 1001–1027 (2021)
12. Kupper, M., Zapata, J.M.: Weakly maxitive set functions and their possibility distributions. Fuzzy Sets Syst. **467**, 108506 (2023)
13. Lacker, D.: A non-exponential extension of Sanov's theorem via convex duality. Adv. Appl. Probab. **52**, 61–101 (2020)
14. Puhalskii, A.: In: Large Deviations and Idempotent Probability. Monographs and Surveys in Pure and Applied Mathematics, vol. 119, Chapman & Hall/CRC Boca Raton, FL (2001)
15. Shilkret, N.: Maxitive measure and integration. Indag. Math. **33**, 109–116 (1971)
16. Tan, Y., Zong, G.: Large deviation principle for random variables under sublinear expectations on $\mathbb{R}^d$. J. Math. Anal. Appl. **488**(2), 124110 (2020)
17. Varadhan, S.R.S.: Asymptotic probabilities and differential equations. Commun. Pure Appl. Math. **19**, 261–286 (1966)
18. Walley, P.: In: Statistical Reasoning with Imprecise Probabilities. Monographs on Statistics and Applied Probability, vol. 42, Chapman and Hall, Ltd., London (1991)
19. Yan, J.: Deviations and asymptotic behavior of convex and coherent entropic risk measures for compound Poisson process influenced by jump times. Stat. Probab. Lett. **125**, 71–79 (2017)
20. Zapata, J.M.: Representation of weakly maxitive monetary risk measures and their rate functions. J. Math. Anal. Appl. **524**(1), 127072 (2023)

Author Index

A
Álvarez-Esteban, Pedro C. 1
Ansari, Jonathan 9, 234

B
Baz, Juan 18
Biernacki, Christophe 27
Blessing, Jonas 36
Broomell, Stephen B. 44
Bukovšek, Damjana Kokol 334

C
Calcagnì, Antonio 59
Campoy-Nieves, Alejandro 273
Castronovo, Lydia 67
Cena, Anna 225
Cheng, Chi-Bin 241
Colubi, Ana 217

D
D'Urso, Pierpaolo 126
de Angelis, Marco 76
De Giovanni, Livia 126
Di Lascio, F. Marta L. 85
Di Perna, Vincenzo 93
Díaz, Irene 18
Dietrich, Nicolas Pascal 102
Dietrich, Stefan 110
Donzé, Laurent 405
Dorfinger, Peter 363
Durante, Fabrizio 118

E
Efromovich, Sam 134
Eichler, Annika 175

F
Federico, Lorenzo 126
Ferraro, Maria Brigida 93
Filzmoser, Peter 325, 357
Fontana, Roberto 349
Fritz, Tobias 142
Fuchs, Sebastian 234, 249, 548
Fuksman, Lirit 134

G
García-Escudero, Luis A. 1
García-García, José 149
Gerber, Alexander 457
Gil, María Ángeles 149, 466
Gómez-Romero, Juan 273
González-Rodríguez, Gil 158
Graña-Colubi, José 158
Grzegorzewski, Przemyslaw 59, 166
Guerrier, Stephane 532

H
Herlich, Matthias 363
Hirlaender, Simon 175, 363, 449
Hryniewicz, Olgierd 184
Huber, Stefan 449

J
Jansen, Christoph 110
Jiroušek, Radim 192, 500
Jurečková, Jana 200

K
Kaczmarek-Majer, Katarzyna 184
Kain, Verena 175
Kaiser, Henrik 208
Kaiser, Jan 175

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
J. Ansari et al. (Eds.): SMPS 2024, AISC 1458, pp. 563–565, 2024.
https://doi.org/10.1007/978-3-031-65993-5

Kalina, Jan 200
Kamkongkaew, Kanyawee 264
Kontoghiorghes, Louisa 217
Kooakachai, Monchai 264
Košir, Tomaž 334
Kozak, Anna 166
Kratochvíl, Václav 192
Kuntur, Soveatin 225
Kupper, Michael 36
Kwitt, Roland 363

L
Langthaler, Patrick B. 234
Lee, Jo-Huan 241
Limbach, Carsten 249
Lubiano, María Asunción 149
Lukas, Lamminger 175
Lütkebohmert, Eva 9

M
Maislinger, Lea 256
Makapawee, Pitchayanin 264
Manjavacas, Antonio 273
Markatou, Marianthi 281
Mayo-Iscar, Agustín 1
McNicholas, Paul D. 317
Miranda, Enrique 421
Mojškerc, Blaž 334
Molinari, Roberto C. 532
Molina-Solana, Miguel 273
Montes, Ignacio 286, 396, 421
Montes, Susana 18
Mordant, Gilles 294
Müller, Alfred 302, 309

N
Neal, Mackenzie R. 317
Neubert, Steffen 457
Nordhausen, Klaus 371

O
Oguamalam, Jeremy 325
Omladič, Matjaž 334

P
Palacio Vega, Sergio 466
Pappadà, Roberta 85, 118
Pérez-Fernández, Raúl 396
Pergamenchtchikov, Serguei 342
Perrone, Elisa 349
Pfeiffer, Pia 357
Picek, Jan 200
Pochaba, Sabrina 175, 363

R
Radojičić, Una 325, 371
Ramos-Guajardo, Ana Belén 158
Ranalli, Monia 380
Rapallo, Fabio 349
Raymaekers, Jakob 388
Rehrl, Jakob 449
Rocci, Roberto 380
Rockel, Marcus 9
Rodemann, Julian 110
Ros, Julián 396
Rosset, Julien 405
Roy, Priyanka 413

S
Salamanca, Juan J. 421
Saminger-Platz, Susanne 413, 429
Sanfilippo, Giuseppe 67
Santamaria Garcia, Andrea 175
Santiwipanont, Tippawan 440
Scarsini, Marco 309
Schäfer, Georg 449
Schirl, Max 449
Schreyer, Manuela 457
Sinova, Beatriz 466, 474
Škulj, Damjan 483
Stopar, Nik 334
Stromme, Austin J. 491
Sumetkijakan, Songkiat 440
Švadlenka, Radek 500

T
Tenzin, Roman 342
Terán, Pedro 142, 509
Todorov, Valentin 518
Trutschnig, Wolfgang 102, 234, 256, 457
Tschimpke, Marco 457

V
Van Aelst, Stefan 474
Van Keilegom, Ingrid 526
Vandewalle, Vincent 27

Victoria-Feser, Maria-Pia 532
Visonà, Tommaso 540
Vitale, Vincenzina 126

W
Wagner, Christian 44
Wang, Yuping 548

X
Xu, Chenran 175

Y
Yanpaisan, Noppawit 440

Z
Zapata, José M. 554