Biometrika (2018), **105**, 3, *pp.* 547–562
Printed in Great Britain

doi: 10.1093/biomet/asy021
Advance Access publication 15 June 2018

Symmetric rank covariances: a generalized framework for nonparametric measures of dependence

By L. WEIHS, M. DRTON
Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195, U.S.A.
lucaw@uw.edu md5@uw.edu

and N. MEINSHAUSEN
Seminar for Statistics, Eidgenössische Technische Hochschule Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
meinshausen@stat.math.ethz.ch

Summary

The need to test whether two random vectors are independent has spawned many competing measures of dependence. We focus on nonparametric measures that are invariant under strictly increasing transformations, such as Kendall's tau, Hoeffding's D, and the Bergsma–Dassios sign covariance. Each exhibits symmetries that are not readily apparent from their definitions. Making these symmetries explicit, we define a new class of multivariate nonparametric measures of dependence that we call symmetric rank covariances. This new class generalizes the above measures and leads naturally to multivariate extensions of the Bergsma–Dassios sign covariance. Symmetric rank covariances may be estimated unbiasedly using U-statistics, for which we prove results on computational efficiency and large-sample behaviour. The algorithms we develop for their computation include, to the best of our knowledge, the first efficient algorithms for Hoeffding's D statistic in the multivariate setting.

Some key words: Dependence; Hoeffding's D; Independence testing; Kendall's tau; U-statistic.

1. Introduction

Many applications require quantification of the dependence between collections of random variables. Letting $X = (X_1, \ldots, X_r)$ and $Y = (Y_1, \ldots, Y_s)$ be random vectors, we are interested in measures of dependence μ which exhibit the following three properties.

Property 1 (I-consistency). If X and Y are independent, then $\mu(X, Y) = 0$.

Property 2 (D-consistency). If X and Y are dependent, then $\mu(X, Y) \neq 0$.

Property 3 (Monotonic invariance). If $f_1, \ldots, f_r, g_1, \ldots, g_s$ are strictly increasing functions, then $\mu(X, Y) = \mu[\{f_1(X_1), \ldots, f_r(X_r)\}, \{g_1(Y_1), \ldots, g_s(Y_s)\}]$. We also refer to this property as μ being nonparametric.

If μ is I-consistent, then tests of independence can be based on the null hypothesis $\mu(X, Y) = 0$. If μ is also D-consistent, then tests based on consistent estimators of μ are guaranteed to asymptotically reject independence when it fails to hold. When μ is both I- and D-consistent,

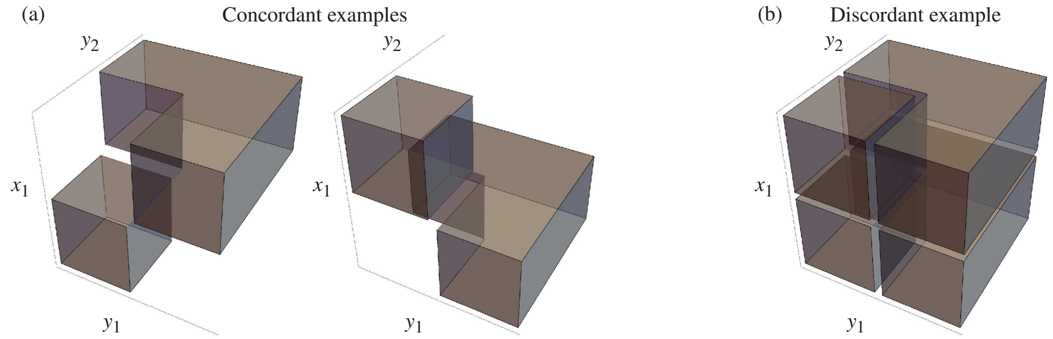

Fig. 1. The bivariate sign covariance τ^* can be defined in terms of the probability of concordance and discordance of four points in $\mathbb{R}^2$ (Bergsma & Dassios, 2014, Fig. 3). Our multivariate extension τ_P^* is based on higher-dimensional generalizations of concordance and discordance. For illustration, let $x^1, \ldots, x^4 \in \mathbb{R}$ and $y^1, \ldots, y^4 \in \mathbb{R}^2$. Considering either of the concordant examples in panel (a), if precisely two tuples (x^i, y^i) fall in each of the two grey regions, then the four tuples are concordant for τ_P^*, but other types of concordance exist. Considering the discordant example in panel (b), if exactly one (x^i, y^i) lies in each of the grey regions, here the two partially obscured regions with smaller x_1 value are just translated copies of the two top regions, then the four tuples are discordant; again, other types of discordance exist. Unlike in the bivariate case, points may be simultaneously concordant and discordant with respect to τ_P^*.

we will simply call it consistent. Monotonic invariance is the intuitive requirement that the level of dependence between two random vectors be invariant under monotonic transformations of any coordinate. Unfortunately, many popular measures of dependence fail to satisfy all of these properties. For instance, Kendall's τ (Kendall, 1938) and Spearman's ρ (Spearman, 1904) are nonparametric and I-consistent but not D-consistent, while the distance correlation (Székely et al., 2007) is consistent but not nonparametric in the above sense.

For bivariate observations, Hoeffding (1948) introduced a nonparametric dependence measure that is consistent for a large class of continuous distributions. Let (X, Y) be a random vector taking values in $\mathbb{R}^2$, with joint and marginal distribution functions F_{XY}, F_X and F_Y. Then the statistic now called Hoeffding's D is defined as $D = \int_{\mathbb{R}^2} \{F_{XY}(x,y) - F_X(x)F_Y(y)\}^2 \, \mathrm{d}F_{XY}(x,y)$. Bergsma & Dassios (2014) introduced a new bivariate dependence measure τ^* that is nonparametric and improves upon Hoeffding's D by guaranteeing consistency for all bivariate mixtures of continuous and discrete distributions. As its name suggests, τ^* generalizes Kendall's τ; where τ counts concordant and discordant pairs of points, τ^* counts concordant and discordant quadruples of points. The proof of consistency of τ^* is considerably more involved than that for D.

Both D and τ^* exhibit symmetries that are obfuscated by their usual definitions. Indeed, as will be made precise, D and τ^* can each be represented as the covariance between signed sums of indicator functions acted on by the subgroup $H = \langle (1\ 4), (2\ 3) \rangle$ of the symmetric group on four elements. We generalize this observation to define a new class of dependence measures called symmetric rank covariances. All such measures are I-consistent and nonparametric, and they include D, τ^*, τ and ρ as special cases. Moreover, our new measures include natural multivariate extensions of τ^* which themselves inspire new notions of concordance and discordance in higher dimensions; see Fig. 1. While symmetric rank covariances need not be D-consistent, we identify a subcollection of measures that are. These consistent measures can be interpreted as testing independence by applying possibly infinitely many independence tests to discretizations of (X, Y). Symmetric rank covariances can readily be estimated using U-statistics, and we show that the use of efficient data structures for orthogonal range queries can give substantial savings. Moreover, we show that under independence, many of the resulting U-statistics are degenerate of order two, thus having non-Gaussian limiting distributions. Most of the proofs are presented in the Supplementary Material.

While our work can be seen as an extension and generalization of classical rank statistics for measuring association, recent interest in dependence testing has produced a variety of approaches to the problem. Broadly, these alternative measures can be organized into those that are based on information theory (Kraskov et al., 2004; Kinney & Atwal, 2014; Reshef et al., 2011, 2016), characteristic functions (Kankainen & Ushakov, 1998; Székely et al., 2007; Hušková & Meintanis, 2008; Rizzo & Szekely, 2016; Böttcher et al., 2017), grid/binning procedures (Heller et al., 2013, 2016; Ma & Mao, 2017), reproducing kernel Hilbert spaces (Gretton et al., 2005), conditional distribution functions (Cui et al., 2015), and generalization or modification of Pearson's correlation coefficient (Breiman & Friedman, 1985; Wang et al., 2017; Zhu et al., 2017). While each of these measures has distinct advantages, our experiments show that symmetric rank covariances are competitive in a number of regimes while remaining simple to state and interpret.

2. Preliminaries

2·1. *Manipulating random and fixed vectors*

We begin by establishing conventions and notation to be used throughout the paper. Let

$$(Z_1, \ldots, Z_{r+s}) = Z = (X, Y) = (X_1, \ldots, X_r, Y_1, \ldots, Y_s)$$

be a random vector taking values in $\mathbb{R}^{r+s}$, and let $(X^i, Y^i) = Z^i$ for $i \in \mathbb{Z}_{>0}$ be a sequence of independent and identically distributed copies of Z. When X and Y are independent we write $X \perp\!\!\!\perp Y$; otherwise we write $X \not\perp\!\!\!\perp Y$. We let F_{XY}, F_X and F_Y denote the cumulative distribution functions for (X, Y), X and Y, respectively.

We will require succinct notation to describe tuples of vectors, possibly permuted. For any $n \geqslant 1$, define $[n] = \{1, \ldots, n\}$. Let $w^1, \ldots, w^n \in \mathbb{R}^d$. Then for $i_1, \ldots, i_m, j_1, \ldots, j_k \in [n]$, let

$$w^{i_1,\ldots,i_m} = w^{(i_1,\ldots,i_m)} = (w^{i_1}, \ldots, w^{i_m}),$$
$$(w^{i_1,\ldots,i_m}, w^{j_1,\ldots,j_k}) = (w^{i_1}, \ldots, w^{i_m}, w^{j_1}, \ldots, w^{j_k}).$$

If $[n]$ appears in the superscript of a vector, it should be interpreted as an ordered vector; that is, we let $w^{[n]} = w^{(1,\ldots,n)} = (w^1, \ldots, w^n)$.

Let S_n be the symmetric group. For $\sigma \in S_n$ and $w^{[n]} \in \mathbb{R}^{d\times n}$, let

$$\sigma w^{[n]} = (w^{\sigma^{-1}(1)}, \ldots, w^{\sigma^{-1}(n)}).$$

This defines a left group action of S_n on $\mathbb{R}^{d\times n}$, which we will often encounter. As our convention is that $[n]$ is an ordered tuple when in a superscript, we have that $\sigma w^{[n]} = w^{\sigma[n]}$ for all $w^{[n]} \in \mathbb{R}^{d\times n}$. We stress that $\sigma(1, \ldots, n) = \{\sigma^{-1}(1), \ldots, \sigma^{-1}(n)\} \neq \{\sigma(1), \ldots, \sigma(n)\}$ in general.

2·2. *Hoeffding's D*

A multivariate version of Hoeffding's D is naturally defined by letting

$$D(X, Y) = \int_{\mathbb{R}^r\times\mathbb{R}^s} \{F_{XY}(x, y) - F_X(x)F_Y(y)\}^2 \,\mathrm{d}F_{XY}(x, y).$$

Since $X \perp\!\!\!\perp Y$ if and only if $F_{XY}(x, y) = F_X(x)F_Y(y)$ for all x and y, it is clear that $X \perp\!\!\!\perp Y$ implies $D(X, Y) = 0$. The converse need not be true, as the next example shows.

Example 1. Let $Z = (X, Y)$ be a bivariate distribution with $\text{pr}(X = 1, Y = 0) = \text{pr}(X = 0, Y = 1) = 1/2$. Then clearly X and Y are not independent but

$$\begin{aligned} D(X, Y) &= \frac{1}{2}\{F_{XY}(1, 0) - F_X(1)F_Y(0)\}^2 + \frac{1}{2}\{F_{XY}(0, 1) - F_X(0)F_Y(1)\}^2 \\ &= \frac{1}{2}(1/2 - 1/2)^2 + \frac{1}{2}(1/2 - 1/2)^2 \;=\; 0. \end{aligned}$$

Thus, $D(X, Y)$ is I-consistent but not D-consistent in general. It is, however, consistent for a large class of continuous distributions.

THEOREM 1 (Multivariate version of Theorem 3.1 in Hoeffding, 1948). *Suppose that X and Y have a continuous joint density f_{XY} and continuous marginal densities f_X and f_Y. Then $D(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$.*

Proof. The bivariate case is treated in Theorem 3.1 in Hoeffding (1948). The proof of the multivariate case is analogous. □

Example 1 highlights that the failure of $D(X, Y)$ to detect all dependence structures can be attributed to the measure of integration $\mathrm{d}F_{XY}$. This suggests the following modification of D, which we call Hoeffding's R:

$$R(X, Y) = \int_{\mathbb{R}^{r+s}} \{F_{XY}(x, y) - F_X(x)F_Y(y)\}^2 \prod_{i=1}^{r} \mathrm{d}F_{X_i}(x_i) \prod_{j=1}^{s} \mathrm{d}F_{Y_j}(y_j).$$

We suspect that it is well known that R is consistent, but we could not find a compelling reference for this fact. For completeness we include a proof in the Supplementary Material.

THEOREM 2. *Let (X, Y) be drawn from a multivariate distribution on $\mathbb{R}^r \times \mathbb{R}^s$. Then $R(X, Y) \geqslant 0$, and $R(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$.*

2·3. *The Bergsma–Dassios sign covariance τ^**

Bergsma & Dassios (2014) defined τ^* only for bivariate distributions, so let $r = s = 1$ for this subsection. While τ^* has a natural definition in terms of concordant and discordant quadruples of points, we will give an alternative definition that will be more useful for our purposes. First, for any $w^{[4]} \in \mathbb{R}^4$ let $I_{\tau^*}(w^{[4]}) = 1_{(w^1, w^2 < w^3, w^4)}$, where $w^1, w^2 < w^3, w^4$ if and only if $\max(w^1, w^2) < \min(w^3, w^4)$. Then, as shown by Bergsma & Dassios (2014),

$$\begin{aligned} \tau^*(X, Y) = E\Big[&\big\{I_{\tau^*}(X^{[4]}) + I_{\tau^*}(X^{4,3,2,1}) - I_{\tau^*}(X^{1,3,2,4}) - I_{\tau^*}(X^{4,2,3,1})\big\} \\ &\times \big\{I_{\tau^*}(Y^{[4]}) + I_{\tau^*}(Y^{4,3,2,1}) - I_{\tau^*}(Y^{1,3,2,4}) - I_{\tau^*}(Y^{4,2,3,1})\big\}\Big]. \end{aligned}$$

Although Bergsma & Dassios (2014) conjectured that τ^* is consistent for all bivariate distributions, the proof of this statement remains elusive. The current understanding of the consistency of τ^* is summarized by the following theorem.

THEOREM 3 (Theorem 1 of Bergsma & Dassios, 2014). *Suppose that (X, Y) are drawn from a bivariate continuous distribution, discrete distribution, or mixture of a continuous and a discrete distribution. Then $\tau^*(X, Y) \geqslant 0$, and $\tau^*(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$.*

Theorem 3 does not apply to any singular distributions; for instance, it is not guaranteed that $\tau^* > 0$ when (X, Y) are generated uniformly on the unit circle in $\mathbb{R}^2$.

3. Symmetric rank covariance

3·1. *Definition and examples*

We now introduce a new class of nonparametric dependence measures that depend on X and Y only through their joint ranks.

Definition 1. *Let $w^{[m]} \in \mathbb{R}^{d\times m}$. Then the joint rank matrix of $w^{[m]}$ is the $[m]$-valued $d \times m$ matrix with (i,j) entry*

$$\mathcal{R}(w^{[m]})_{ij} = 1 + \sum_{k=1}^{m} 1_{(w_i^k < w_i^j)};$$

that is, $\mathcal{R}(w^{[m]})_{ij}$ is the rank of w_i^j among $w_i^1, \ldots, w_i^m$ for $i \in [d]$.

Definition 2. *A rank indicator function of order m and dimension d is a function $I : \mathbb{R}^{d\times m} \to \{0, 1\}$ such that $I\{\mathcal{R}(w^{[m]})\} = I(w^{[m]})$ for all $w^{[m]} \in \mathbb{R}^{d\times m}$. In other words, I depends on its arguments only through their joint ranks.*

Definition 3. *Let I_X and I_Y be rank indicator functions that have equal order m and are of dimensions r and s, respectively. Let H be a subgroup of the symmetric group S_m with an equal number of even and odd permutations. Define*

$$\mu_{I_X,I_Y,H}(X, Y) = E\left[\left\{\sum_{\sigma\in H} \mathrm{sign}(\sigma)\, I_X(X^{\sigma[m]})\right\}\left\{\sum_{\sigma\in H} \mathrm{sign}(\sigma)\, I_Y(Y^{\sigma[m]})\right\}\right]. \quad (1)$$

Then a measure of dependence μ is a symmetric rank covariance if there exist a scalar $c > 0$ and a triple (I_X, I_Y, H) as specified above such that $\mu = c\,\mu_{I_X,I_Y,H}$. More generally, μ is a summed symmetric rank covariance if it is the sum of several symmetric rank covariances.

Some of the symmetric rank covariances we consider have the two rank indicator functions equal to each other, so $I_X = I_Y = I$. In this case, we also use the abbreviation $\mu_{I,H} = \mu_{I,I,H}$.

Remark 1. In Definition 3, the restrictions on H both allow us to generalize several existing nonparametric measures of dependence and afford us a number of general properties that we leverage in our proofs. An interesting alternative choice, suggested by a referee, is to let H be an arbitrary subset of S_m with some partial order $\preceq$ for which H has a unique upper bound $\hat{1}$ and a unique lower bound $\hat{0}$. Upon replacing the sign function in (1) by the function $\sigma \mapsto m(\sigma, \hat{1})$, where $m : S_m \to \mathbb{R}$ is the Möbius function corresponding to $(H, \preceq)$, one obtains a new collection of measures, which one might naturally call Möbius rank covariances, having many of the same properties as the symmetric rank covariances. Indeed, all special cases of symmetric rank covariances considered in this paper are also Möbius rank covariances, and one may recover Proposition 2 for these measures. While the study of such measures is beyond the scope of this paper, an avenue of future research might be to investigate the properties of such measures when using the lattice structure on S_m from Duquenne & Cherfouh (1994).

Remark 2. Recall from § 2·3 that for any $z^{[4]} \in \mathbb{R}^4$ we write $z^1, z^2 < z^3, z^4$ to mean $\max(z^1, z^2) < \min(z^3, z^4)$. To simplify the definitions of rank indicator functions, we generalize this notation as follows. Let $\sim$ be any binary relation on $\mathbb{R}^d$. Then for $z^{[l]} \in \mathbb{R}^{d\times l}$ and $w^{[k]} \in \mathbb{R}^{d\times k}$ we write $z^1, \ldots, z^l \sim w^1, \ldots, w^k$ to mean $z^i \sim w^j$ for all $(i, j) \in [l] \times [k]$.

It is easy to show that many existing nonparametric measures of dependence are symmetric rank covariances.

PROPOSITION 1. *Let X and Y take values in $\mathbb{R}^r$ and $\mathbb{R}^s$, respectively. Consider the permutation groups $H_\tau = \langle(1\ 2)\rangle$ and $H_{\tau^*} = \langle(1\ 4), (2\ 3)\rangle$.*

(i) *Bivariate case ($r = s = 1$): Kendall's τ, its square τ^2, and τ^* of Bergsma & Dassios (2014) are symmetric rank covariances. Specifically,*

$$\tau = \mu_{I_\tau, H_\tau}, \quad \tau^2 = \mu_{I_{\tau^2}, H_{\tau^*}}, \quad \tau^* = \mu_{I_{\tau^*}, H_{\tau^*}},$$

where the one-dimensional rank indicator functions are defined as

$$I_{\tau^*}(w^{[4]}) = 1_{(w^1, w^2 < w^3, w^4)}, \quad I_\tau(w^{[2]}) = 1_{(w^1 < w^2)}, \quad I_{\tau^2}(w^{[4]}) = I_\tau(w^{1,4}) I_\tau(w^{2,3}).$$

(ii) *General case ($r, s \geqslant 1$): both D and R are symmetric rank covariances. Specifically,*

$$D = \frac{1}{4}\mu_{I_{D,r}, I_{D,s}, H_{\tau^*}}, \quad R = \frac{1}{4}\mu_{I_{R,r,s,X}, I_{R,r,s,Y}, H_{\tau^*}},$$

where for any $d \geqslant 1$ we define

$$I_{D,d}(w^{[5]}) = 1_{(w^1, w^2 \preceq w^5)} 1_{(w^3, w^4 \npreceq w^5)} \quad (w^1, \ldots \in \mathbb{R}^d),$$
$$I_{R,r,s,X}(w^{[4+r+s]}) = 1_{\{w^1, w^2 \preceq (w_1^5, \ldots, w_r^{4+r})^{\mathrm{T}}\}} 1_{\{w^3, w^4 \npreceq (w_1^5, \ldots, w_r^{4+r})^{\mathrm{T}}\}} \quad (w^1, \ldots \in \mathbb{R}^r),$$
$$I_{R,r,s,Y}(w^{[4+r+s]}) = 1_{\{w^1, w^2 \preceq (w_1^{5+r}, \ldots, w_s^{4+r+s})^{\mathrm{T}}\}} 1_{\{w^3, w^4 \npreceq (w_1^{5+r}, \ldots, w_s^{4+r+s})^{\mathrm{T}}\}} \quad (w^1, \ldots \in \mathbb{R}^s),$$

with $w^i \preceq w^j$ if and only if $w^i_\ell \leqslant w^j_\ell$ for all $\ell \in [d]$.

Remark 3. In Proposition 1 we see that, unlike for $I_{D,d}$, the length of the input tuples to $I_{R,r,s,X}$ and $I_{R,r,s,Y}$ grows with r and s. While this may seem surprising, it is an immediate consequence of the fact that R integrates against the product measure $\prod_{i=1}^r \mathrm{d}F_{X_i}(x_i) \prod_{j=1}^s \mathrm{d}F_{Y_j}(y_j)$, each component of which requires its own independent observation.

Remark 4. The bivariate dependence measure Spearman's ρ can be written as

$$\rho(X, Y) = 6E\big[1_{(X^1 < X^2 < X^3)} \{1_{(Y^1 < Y^2 < Y^3)} + 1_{(Y^1 < Y^3 < Y^2)} + 1_{(Y^2 < Y^1 < Y^3)} - 1_{(Y^3 < Y^1 < Y^2)} - 1_{(Y^2 < Y^3 < Y^1)} - 1_{(Y^3 < Y^2 < Y^1)}\}\big]. \quad (2)$$

In light of Lemma 1, one might expect ρ to be a symmetric rank covariance. However, upon examining which of the indicators are negated in (2), one sees that the permutations do not respect the sign operation of the permutation group S_3. For instance, $1_{(Y^1 < Y^2 < Y^3)}$ and $1_{(Y^1 < Y^3 < Y^2)}$ are related through a single transposition and yet the terms have the same sign above. While it seems difficult to prove conclusively that ρ is not a symmetric rank covariance, this observation suggests that it is not. Somewhat surprisingly, however, ρ is a summed symmetric rank covariance, which can be seen by expressing ρ as

$$\rho(X,Y) = 3\,E\big\{b(X^{[3]})b(Y^{[3]}) + b(X^{[3]})b(Y^{1,3,2}) + b(X^{[3]})b(Y^{2,1,3})\big\},$$

where $b(z^{[3]}) = 1_{(z^1<z^2<z^3)} - 1_{(z^3<z^2<z^1)}$ for all $z^i \in \mathbb{R}$.

3·2. *General properties*

Although many interesting properties of symmetric rank covariances depend on the choice of the group H and of the indicators I_X and I_Y, several properties hold for all such choices.

PROPOSITION 2. *Let μ be a symmetric rank covariance. Then μ is nonparametric and I-consistent. If ν is another symmetric rank covariance, then so is the product $\mu\nu$.*

Remark 5. While Proposition 2 guarantees that all symmetric rank covariances are nonparametric and I-consistent, showing that such measures are D-consistent must be done on a case-by-case basis and can, in general, be difficult. In the Supplementary Material we produce, through a natural generalization of Hoeffding's D and R, a collection of summed symmetric rank covariances for which proving D-consistency is relatively straightforward. Thus the Supplementary Material serves to demonstrate one strategy by which we may recover D-consistency while also generating a collection of candidate measures for further study.

The property for products in particular justifies squaring symmetric rank covariances, as was done for bivariate rank correlations in Leung & Drton (2018). Later, it will be useful to express symmetric rank covariances in an equivalent form.

LEMMA 1. *In reference to* (1), *we have*

$$\mu_{I_X,I_Y,H}(X,Y) = |H|\,E\left\{I_X(X^{[m]}) \sum_{\sigma\in H} \mathrm{sign}(\sigma)\, I_Y(Y^{\sigma[m]})\right\} \qquad (3)$$

$$= |H|\,E\left\{I_Y(Y^{[m]}) \sum_{\sigma\in H} \mathrm{sign}(\sigma)\, I_X(X^{\sigma[m]})\right\}. \qquad (4)$$

4. MULTIVARIATE τ^*

Recall from Proposition 1 that $\tau^* = \mu_{I_{\tau^*},H_{\tau^*}}$. Multivariate extensions of τ^* should simultaneously capture essential characteristics and permit enough flexibility to define interesting measures of high-order dependence. As a first step to distilling these essential characteristics, it seems natural that any multivariate extension of τ^* should use the same permutation subgroup H_{τ^*}.

Remark 6. There are 30 distinct subgroups of S_4, exactly 20 of which have an equal number of even and odd permutations and thus could be used in the definition of a symmetric rank covariance. Given these many possible choices, it may seem surprising that H_{τ^*} appears in the definition of so many existing measures of dependence, including τ^*, τ^2, D and R. Some intuition as to the ubiquity of H_{τ^*} can be gleaned from the proof of Proposition 1; see the Supplementary Material, where we show that H_{τ^*} arises naturally from an expansion of $\{F_{XY}(x,y) - F_X(x)F_Y(y)\}^2$.

It now remains to find an appropriate generalization of I_{τ^*}. To better characterize I_{τ^*}, we require the following definition.

DEFINITION 4. *Let I be a rank indicator function of order m and dimension d. The permutations $\sigma \in S_m$ such that $I(\sigma w^{[m]}) = I(w^{[m]})$ for all $w^{[m]} \in \mathbb{R}^{d\times m}$ form a group which we refer to as the invariance group G of I. For any symmetric rank covariance $\mu_{I_X,I_Y,H}$, let G_X and G_Y be the invariance groups of I_X and I_Y, respectively. We then call $G = G_X \cap G_Y$ the invariance group of $\mu_{I_X,I_Y,H}$.*

We now single out two properties of I_{τ^*}:

Property 4. I_{τ^*} is a rank indicator function of order 4;

Property 5. the invariance group of I_{τ^*} is $\langle(1\ 2), (3\ 4)\rangle$.

These properties inspire the following definition.

DEFINITION 5. *A symmetric rank covariance $\mu_{I_X,I_Y,H}$ is a τ^* extension if I_X and I_Y are rank indicators of order 4 with invariance group $\langle(1\ 2), (3\ 4)\rangle$ and $H = H_{\tau^*}$.*

From the possible τ^* extensions we consider two notable candidates.

DEFINITION 6. *For any $d \geqslant 1$, let $I_P : \mathbb{R}^{d\times 4} \to \{0, 1\}$ be the rank indicator, where for any $w^{[4]} \in \mathbb{R}^{d\times 4}$ we have $I_P(w^{[4]}) = 1_{(w^3,w^4 \not\preceq w^1,w^2)}$. We then call $\mu_{I_P,I_P,H_{\tau^*}}$ the multivariate partial τ^* and write $\tau^*_P = \mu_{I_P,I_P,H_{\tau^*}}$.*

The definition of I_P is inspired by I_D; see Proposition 1.

DEFINITION 7. *For any $d \geqslant 1$, let $I_J : \mathbb{R}^{d\times 4} \to \{0, 1\}$ be the rank indicator, where for any $w^{[4]} \in \mathbb{R}^{d\times 4}$ we have $I_J(w^{[4]}) = 1_{(w^1,w^2 \prec w^3,w^4)}$. We then call $\mu_{I_J,I_J,H_{\tau^*}}$ the multivariate joint τ^* and write $\tau^*_J = \mu_{I_J,I_J,H_{\tau^*}}$.*

Our definition of τ^*_J comes immediately from τ^* upon replacing the total order $<$ with $\prec$. Although this might be the most intuitive multivariate extension of τ^*, it is easily seen to not be D-consistent, as the next example shows. In both of the above definitions, the extensions reduce to τ^* when $r = s = 1$.

Example 2. Let $X = (X_1, \ldots, X_r)$, where r is even and $X_1, \ldots, X_r \sim \mathrm{Ber}(1/2)$ are independent. Now let $Y = \textsc{xor}(X_1, \ldots, X_r)$, that is, let $Y = 1$ if $\sum_{i=1}^r X^r$ is odd and $Y = 0$ otherwise. Then, letting (X^i, Y^i) be independent and identically distributed replicates of (X, Y), $I_J(X^{[4]}) = 1$ if and only if $X^1 = X^2 = (0, \ldots, 0)$ and $X^3 = X^4 = (1, \ldots, 1)$. Thus $I_j(X^{[4]}) = 1$ implies $Y^1 = Y^2 = Y^3 = Y^4 = 0$ and hence that $\sum_{\sigma\in H_{\tau^*}} \mathrm{sign}(\sigma)\, I_J(Y^{\sigma[4]}) = 0$. Therefore we have that $\tau^*_J(X, Y) = 0$ while $X \not\perp\!\!\!\perp Y$.

This behaviour occurs only when r is even. If r is odd, then $\tau^*_J(X, Y) = 2^{-4r+2}$.

Unlike for τ^*_J, we have yet to discover an example where $\tau^*_P(X, Y)$ is zero when $X \not\perp\!\!\!\perp Y$. This leads us to conjecture that τ^*_P, like the subclass of measures from the Supplementary Material, is D-consistent.

5. Estimation via U-statistics

5·1. *Standard form of U-statistics estimating symmetric rank covariances*

Symmetric rank covariances can be readily estimated using U-statistics. As naïvely computing these U-statistics is often intractable, we will show how efficient data structures from computational geometry can often be used to substantially decrease run-time. We will then consider the asymptotic properties of our estimators and exhibit a collection of symmetric rank covariances whose U-statistics have non-Gaussian asymptotic distributions under the null hypothesis of independence. In the bivariate setting, these results will allow us to show explicitly that the asymptotic distributions of the U-statistics corresponding to D, R and τ^* are, up to scaling, the same when X and Y are continuous and independent, a behaviour first observed by Nandy et al. (2016).

Let $\mu = \mu_{I_X,I_Y,H}$ be as in (1). We let $\kappa : \mathbb{R}^{(r+s)\times m} \to \mathbb{R}$ be the symmetrized kernel function defined by

$$\kappa(z^1, \ldots, z^m) = \frac{1}{m!} \sum_{\sigma \in S_m} k(z^{\sigma[m]}),$$

where the unsymmetrized kernel function $k : \mathbb{R}^{(r+s)\times m} \to \mathbb{R}$ is defined by

$$k(z^{[m]}) = \left\{ \sum_{\sigma \in H} \text{sign}(\sigma)\, I_X(x^{\sigma[m]}) \right\} \left\{ \sum_{\sigma \in H} \text{sign}(\sigma)\, I_Y(y^{\sigma[m]}) \right\}.$$

Then we define, for $n \geqslant m$ and $z^{[n]} \in \mathbb{R}^{d\times n}$,

$$U_\mu(z^{[n]}) = \frac{1}{\binom{n}{m}} \sum_{1\leqslant i_1 < \cdots < i_m \leqslant n} \kappa(z^{i_1,\ldots,i_m}). \tag{5}$$

We call U_μ the U-statistic corresponding to μ. Clearly, $U_\mu(Z^{[n]})$ is unbiased for $\mu(X, Y)$. For ease of computation we will sometimes rewrite κ using the following proposition.

Proposition 3. *For any $z^{[m]} \in \mathbb{R}^{d\times m}$,*

$$\kappa(z^{[m]}) = \frac{|H|}{m!} \sum_{\gamma \in S_m} I_X(x^{\gamma[m]}) \sum_{\sigma \in H} \text{sign}(\sigma)\, I_Y(y^{\sigma\gamma[m]}) \tag{6}$$

$$= \frac{|H|}{m!} \sum_{\gamma \in S_m} I_Y(y^{\gamma[m]}) \sum_{\sigma \in H} \text{sign}(\sigma)\, I_X(X^{\sigma\gamma[m]}). \tag{7}$$

5·2. *Efficient computation*

The U-statistics defined by (5) are a sum over n choose m elements and thus, assuming that the indicator functions I_X and I_Y can be computed in $O(m)$ time, require $O(m\, n^m)$ operations in a naïve computation. While this might be feasible for small m and n, it will be prohibitive for even moderate sample sizes. Although subsampling can be used to approximate our statistics of interest, it is not always clear how many samples of what size should be taken to obtain an acceptable approximation error, and when many such samples are needed, subsampling approximations may not be fast. Fortunately, we show that when specializing to the U-statistics estimating D, R, τ^*_P and

τ_J^*, the use of efficient data structures from computational geometry can reduce the asymptotic run-time of the computations. While our observations do not generalize to all symmetric rank covariances, there appear to be many, such as τ^2, for which a similar approach can be used to reduce run-time. For very large samples, these computational strategies could be combined with subsampling procedures to achieve low approximation error more rapidly.

For the remainder of this section we assume that we have observed data $z^{[n]} \in \mathbb{R}^{d\times n}$. Moreover, to simplify our run-time analyses, we will assume that d is bounded, so that for any functions $f, g : \mathbb{N} \to \mathbb{N}$ and $h : \mathbb{N}^2 \to \mathbb{N}$ we have that $O\{f(d) + g(d)h(n, d)\} = O\{h(n, d)\}$.

As the U-statistics defined above depend on $z^{[n]}$ only through their joint ranks, we will make the further assumption that $z^{[n]} = \mathcal{R}(z^{[n]}) \in [n]^{d\times n}$, so that we have transformed $z^{[n]}$ into its corresponding matrix of joint ranks. The computational effort of this procedure is $O\{n \log_2(n)\}$, and as none of the algorithms we will present has a run-time less than this, performing this pre-processing step does not change the overall analysis.

In the bivariate case, it follows easily from the discussion in Hoeffding (1948) that D can be computed in $O(n \log_2 n)$ time, while more recently it has been shown that τ^* can be computed in $O(n^2)$ time (Heller & Heller, 2016; Weihs et al., 2016). These computational savings largely rely on the ability to efficiently perform orthogonal range queries.

Definition 8. *Let $z^{[n]} \in \mathbb{R}^{d\times n}$. Then the question of how many z^i lie in $B \subset \mathbb{R}^d$ is an orthogonal range query on $\{z^1, \ldots, z^n\}$ if $B = I_1 \times \cdots \times I_d$ and, for $1 \leqslant i \leqslant d$, I_i is an interval of the form (l_i, u_i), $[l_i, u_i)$, $(l_i, u_i]$ or $[l_i, u_i]$ for some $l_i, u_i \in \mathbb{R}$.*

As the next proposition shows, by using a simple dynamic programming approach one can easily construct an n^d tensor so that any orthogonal range query on $z^{[n]}$ can be answered in $O(1)$ time. See Heller & Heller (2016) for the bivariate case.

Proposition 4. *Let $z^{[n]} \in \mathbb{R}^{d\times n}$ be such that $z^{[n]} = \mathcal{R}(z^{[n]})$. Then let $A \in \mathbb{N}^{(n+1)\times\cdots\times(n+1)}$ be a d-dimensional tensor, indexed by elements of $\{0, \ldots, n\}^d$, whose $(i_1, \ldots, i_d) \in \{0, \ldots, n\}^d$ entry is*

$$A(i_1, \ldots, i_d) = \sum_{i=1}^{n} 1_{\{z^i=(i_1,\ldots,i_d)\}},$$

so that $A(i_1, \ldots, i_d)$ equals the number of elements z^i with value $(i_1, \ldots, i_d)$. Now define $B \in \mathbb{N}^{(n+1)\times\cdots\times(n+1)}$ recursively so that it has $(i_1, \ldots, i_d) \in \{0, \ldots, n\}^d$ entry $B(i_1, \ldots, i_d) = 0$ if any $i_j = 0$ and

$$B(i_1, \ldots, i_d) = A(i_1, \ldots, i_d) + \sum_{s=1}^{d} \sum_{\substack{\ell\in\{0,1\}^d\setminus\{0_d\} \\ \sum_k \ell_k = s}} (-1)^{s+1} B\{(i_1, \ldots, i_d) - \ell\}$$

otherwise. Then for any $l = (l_1, \ldots, l_d), u = (u_1, \ldots, u_d) \in \{0, \ldots, n\}^d$, the answer to the orthogonal range query as to how many z^i lie in $B = (l_1, u_1] \times \cdots \times (l_d, u_d]$ is

$$\sum_{\ell\in\{0,1\}^d} (-1)^{\sum_{j=1}^{d} \ell_j} B(l_1^{\ell_1}\, u_1^{1-\ell_1}, \ \ldots, \ l_d^{\ell_d}\, u_d^{1-\ell_d}).$$

When d is bounded and B is given, the above sum takes $O(1)$ time to compute.

Proof. This follows from application of the inclusion-exclusion principle. □

Unfortunately, the above tensor takes $O(n^d)$ time to construct, so when $d \geqslant m$ this procedure already takes at least as long as computing the U-statistic naïvely. In such cases, the range-tree data structure provides a better balance between quickly computing the answer to an orthogonal range query and the effort required for its construction.

PROPOSITION 5 (de Berg et al., 2008). *Let $z^{[n]} \in \mathbb{R}^{d\times n}$. There exists a data structure, called a range-tree, which takes $O\{n\log_2(n)^{d-1}\}$ time to construct and can answer any orthogonal range query on $z^{[n]}$ in $O\{\log_2(n)^{d-1}\}$ time.*

See de Berg et al. (2008, § 5) for a detailed exposition on range-trees, along with a discussion of the above proposition and orthogonal range queries in general. Because to the best of our knowledge there exists no open source, for completely general implementation of range-trees, we make such an implementation freely available at https://github.com/Lucaweihs/range-tree. Range-trees are closely related to binary search trees, such as red-black trees, which have previously been used to efficiently compute the U-statistics corresponding to τ and τ^* (Christensen, 2005; Weihs et al., 2016). Using these data structures, we can achieve substantial run-time savings.

When using the algorithms described in the Supplementary Material, the asymptotic run-times of computing U_D, U_R, $U_{\tau_P^*}$ and $U_{\tau_J^*}$ are $O\{n\log_2(n)^{d-1}\}$, $O(n^d)$, $O\{n^2\log_2(n)^{2d-1}\}$ and $O\{n^2\log_2(n)^{2d-1}\}$, respectively. When computing these statistics naïvely, their asymptotic run-times are $O(n^5)$, $O(n^{4+d})$, $O(n^4)$ and $O(n^4)$, respectively.

5·3. *Null asymptotics*

Determining the asymptotic distribution of U_μ under the null hypothesis of independence, i.e., $X \perp\!\!\!\perp Y$, requires an understanding of the functions

$$\kappa_i(z^1,\ldots,z^i) = E\{\kappa(z^1,\ldots,z^i,Z^{i+1},\ldots,Z^m)\}.$$

To this end, we introduce some simplifying lemmas and propositions.

LEMMA 2. *Suppose that $X \perp\!\!\!\perp Y$. Let $S \subset [m]$ and let G be the invariance group corresponding to μ. Partition H into equivalence classes $E_1,\ldots,E_t$, where $h,h' \in H$ are equivalent if there exists $g \in G$ such that $gh(i) = h'(i)$ for all $i \in S$. If each E_i contains an equal number of even and odd permutations, then for any $z_1,\ldots,z_m \in \mathbb{R}^{r+s}$ we have $E\{k(W^{[m]})\} = 0$, where $W^i = z^i$ if $i \in S$ and $W^i = Z^i$ otherwise.*

Lemma 2 allows us to identify conditions guaranteeing that U_μ is degenerate, that is, cases in which $n^{1/2}(U_\mu - EU_\mu)$ converges to zero in probability.

PROPOSITION 6. *Suppose that the conditions of Lemma 2 hold for μ whenever S is a singleton set. If $X \perp\!\!\!\perp Y$, then $\kappa_1 \equiv 0$ and hence U_μ is a degenerate U-statistic.*

As an application of Lemma 2 and Proposition 6, we deduce the known result that τ^*, D and R are degenerate U-statistics under independence and that their κ_2 functions take a simple form.

LEMMA 3. *Let I_X and I_Y be rank indicators of order $m \geqslant 4$ and dimensions r and s, respectively. Let $\mu = \mu_{I_X,I_Y,H_{\tau^*}}$ be a symmetric rank covariance. Suppose that $X \perp\!\!\!\perp Y$. If the invariance group*

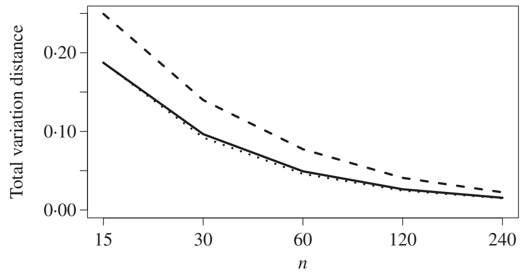

Fig. 2. Total variation distance from kernel density estimators of the finite-sample distributions of $U_{\tau*}$ (solid), U_D (dashed) and U_R (dotted) to the probability density functions of their asymptotic distributions. The horizontal axis is plotted on a log-scale. Here $n \in \{15, 30, 60, 120, 240\}$ is the sample size. The finite-sample distributions are quite close to the asymptotic distributions even when n is only around 60.

of μ contains the subgroup $G = \langle(1\ 2), (3\ 4)\rangle$, then $\kappa_1(z^1) \equiv 0$, so U_μ is a degenerate U-statistic and

$$\kappa_2(z^1, z^2) = \frac{4}{\binom{m}{2}} E\{a_{I_X}(x^1, x^2, X^{3,\ldots,m})\} E\{a_{I_Y}(y^1, y^2, Y^{3,\ldots,m})\},$$

where for any rank indicator I of order $m \geqslant 4$ we define

$$a_I(w^{[m]}) = \sum_{\sigma \in H_{\tau*}} \mathrm{sign}(\sigma) I(w^{\sigma[m]}).$$

By construction, all multivariate τ^* extensions satisfy the above conditions, as do τ^*, D and R.

As noted by Nandy et al. (2016), the U-statistics corresponding to τ^* and D have, up to a scalar multiple, the same asymptotic distribution under the null hypothesis that $X \perp\!\!\!\perp Y$ and (X, Y) are drawn from a continuous bivariate distribution. We give a simple proof of this fact, as well as showing that U_R has an asymptotic distribution that is also a scalar multiple of the others, and clarify the constant multiple by which the distributions differ.

PROPOSITION 7. *Let*

$$Z = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{i^2 j^2} (\chi^2_{1,ij} - 1),$$

where $\{\chi^2_{1,ij} : i, j \in \mathbb{N}_+\}$ is a collection of independent and identically distributed χ^2_1 random variables. Then $n\, U_{\tau*} \to 36Z/\pi^4$, and both $n\, U_D, n\, U_R \to Z/\pi^4$ in distribution.

To better understand at what sample size n the finite-sample distributions of $U_{\tau*}$, U_D and U_R become well approximated by their asymptotic distributions, Fig. 2 plots the total variation distance between kernel density estimates of the finite-sample distributions of $U_{\tau*}$, U_D and U_R for $n \in \{15, 30, 60, 120, 240\}$ against the probability density functions of their asymptotic distributions. We observe good agreement even for $n = 60$. Unfortunately, clarifying the exact asymptotic behaviour in higher dimensions seems significantly more difficult than in the bivariate

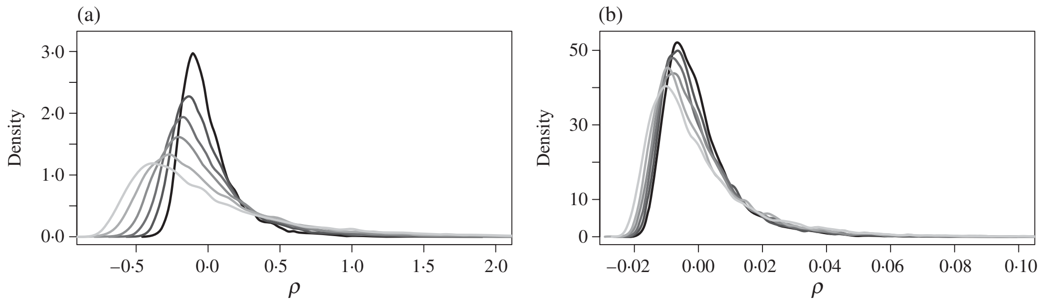

Fig. 3. Kernel density estimates of the finite-sample distributions of (a) $n\,U_{\tau_J^*}$ and (b) $n\,U_D$ for samples of size $n = 70$ taken from (X, Y), where $X, Y_1, Y_2 \sim N(0, 1)$, (Y_1, Y_2) are jointly normal with correlation ρ, and $X \perp\!\!\!\perp Y$. Here ρ varies in $\{0, 1/5, \ldots, 1\}$, with the lighter lines corresponding to kernel density estimates for smaller values of ρ.

case. In part this is because, unlike in the continuous bivariate case, the distributions of the random vectors X and Y influence the asymptotic properties of our multivariate U-statistics. Indeed, even when $r = 1$, $s = 2$, and X and Y are normally distributed, Fig. 3 suggests that the correlation between Y_1 and Y_2 affects the large-sample behaviour. Because of these difficulties, we leave this problem for future work.

6. Simulations

All of the following experiments were run in R (R Development Core Team, 2018) using the package SymRC, which can be obtained from https://github.com/Lucaweihs/SymRC.

We test whether a univariate response Y is independent of a set of covariates $X = (X_1, \ldots, X_r)$, using the U-statistics corresponding to D, R, τ_P^* and τ_J^*. As explicit asymptotic distributions for U_D, U_R, $U_{\tau_P^*}$ and $U_{\tau_J^*}$ are not known, we will use permutation tests. Unfortunately the computational complexities of U_R, $U_{\tau_P^*}$ and $U_{\tau_J^*}$ are such that, while it is possible to perform permutation tests for a single moderately sized sample, it becomes computationally prohibitive to perform the many thousands of tests needed for Monte Carlo approximation of power. We therefore approximate the results of permutation tests in the following way. First we create a reference distribution for our U-statistic of interest under $X \perp\!\!\!\perp Y$ by, for $R = 1000$ replications, randomly generating $x^1, \ldots, x^n$ independently from the marginal distribution of X and $y^1, \ldots, y^n$ independently from the marginal distribution of Y and saving the value of the U-statistic for this dataset. For an independent and identically distributed sample $\mathcal{D} = \{(\bar{x}^1, \bar{y}^1), \ldots, (\bar{x}^n, \bar{y}^n)\}$ from the true joint distribution of (X, Y), we then compute a p-value as the proportion of observations in the reference distribution that are greater than or equal to the value of the U-statistic when computed on $\mathcal{D}$.

This procedure differs from a standard permutation test only in that the reference distribution, and hence the critical value for rejection, is slightly different. Empirical tests using small samples suggest, however, that results obtained with the above procedure generalize well to those obtained using a true permutation test. Computing U_D is sufficiently fast that we do not need to use the above procedure and instead employ a standard permutation test.

For comparison, we compute the power of the permutation test based on the distance covariance d_{cov}, as computed by the R package energy (Rizzo & Szekely, 2016). While d_{cov} is not a nonparametric measure, we can create a nonparametric version of it by instead considering the measure $d_{\mathrm{cov}}^{\mathrm{rank}} = d_{\mathrm{cov}}\{(F_{X_1}[X_1], \ldots F_{X_r}[X_r]), (F_{Y_1}[Y_1], \ldots, F_{Y_s}[Y_s])\}$; we also compute the power of a test based on this measure. Of course, in practice, we do not know $F_{X_1}, \ldots, F_{Y_s}$ and so must estimate them with the marginal empirical cumulative distribution functions. In each of our simulations, we estimate the power using 1000 sample datasets from the relevant joint distribution.

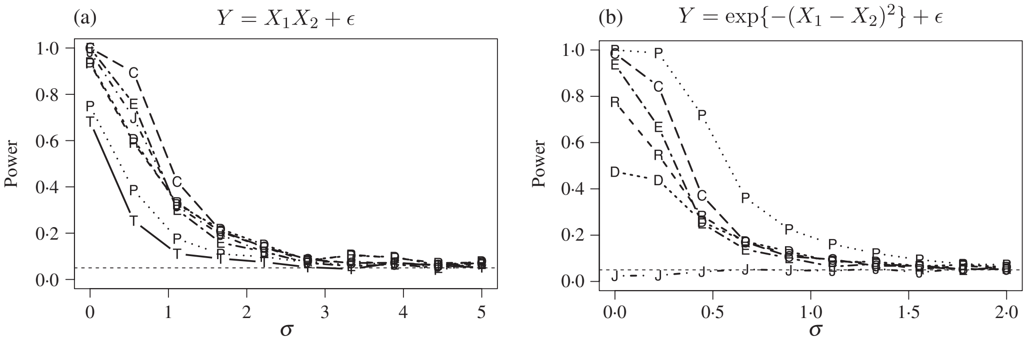

Fig. 4. Empirical power of permutation tests using U_D (short-dashed line with symbol D), U_R (medium-dashed, R), $U_{\tau_P^*}$ (dotted, P), $U_{\tau_J^*}$ (dot-dashed, J), $d_{\rm cov}$ (long-dashed, C), and $d_{\rm cov}^{\rm rank}$ (short-long-dashed, E) in the continuous case. The horizontal dashed line shows the nominal 0·05 level. Here σ is the standard deviation of the additive noise ϵ. In panel (a), the line with symbol T corresponds to $d_{\rm cov}$ after applying the strictly increasing transformation $y \mapsto \text{sign}(y)\log(|y| + 10)$ to Y.

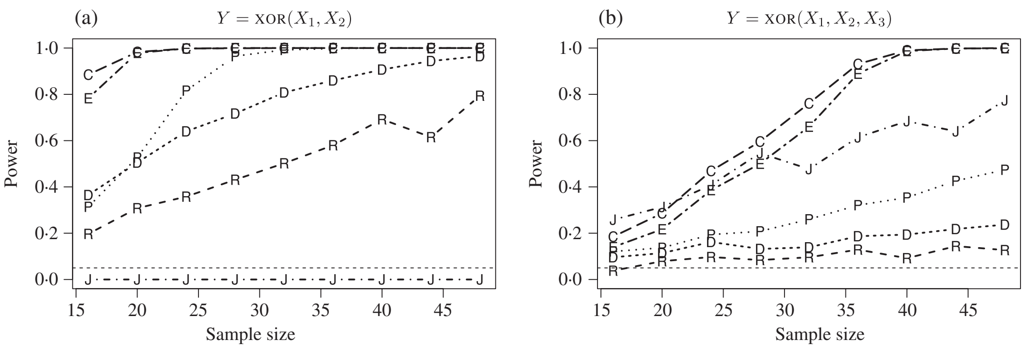

Fig. 5. Empirical power of permutation tests of independence for a jointly discrete distribution when $n \in \{16, 20, \ldots, 48\}$. See the caption of Fig. 4 for the correspondence between line types and tests.

We consider two cases in which we generate samples of size 50 from jointly continuous distributions. First we set $r = 2$, let X_1 and X_2 be independent samples from a $N(0, 1)$ distribution, and take $Y = X_1X_2 + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$ with $\sigma \in \{0, \ldots, 5\}$. Figure 4(a) depicts the power of the tests of the hypothesis that $X \perp\!\!\!\perp Y$. For comparison, we include the power of the distance covariance when Y has been monotonically transformed by the function $f(y) = \text{sign}(y)\log(|y| + 10)$; this transformation substantially reduces the power of the distance covariance while having no impact on the power of the other tests as they are nonparametric. In the second case, we let X_1, X_2 and ϵ be as above but define $Y = \exp\{-(X_1 - X_2)^2\} + \epsilon$. Figure 4(b) displays the power of the tests as σ varies in $[0, 2]$. The power of the test based on τ_J^* is always near the nominal 0·05 level, suggesting that $\tau_J^*(X, Y) = 0$ and hence that τ_J^* is not D-consistent in this case.

We also consider two cases in which (X, Y) is generated from a jointly discrete distribution. Unlike in the continuous case, the sample size n will vary with $n \in \{16, 20, \ldots, 48\}$. First, we set $r = 2$, let X_1 and X_2 be independent samples from a Ber(1/2) distribution, and take $Y = \text{XOR}(X_1, X_2)$. We compute the power of our tests for various sample sizes and plot the results in Fig. 5(a). As we would expect from Example 2, the power of the test based on τ_J^* equals zero at all sample sizes. Secondly, we set $r = 3$, let X_1, X_2 and X_3 be independent samples from a Ber(1/2) distribution, and define $Y = \text{XOR}(X_1, X_2, X_3)$. Figure 5(b) displays the power of the tests in this setting. Unlike in the previous case, the power of the test based on τ_J^* is quite high; again, recall from Example 2 that this is because r is odd.

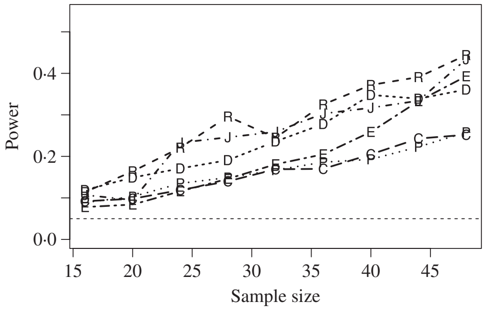

Fig. 6. Empirical power of permutation tests of independence when $n \in \{16, 20, \ldots, 48\}$. Here $Y \sim \text{Ber}[\text{expit}\{6\sin(X_1X_2)\}]$ so that the joint distribution (X, Y) is neither jointly continuous nor jointly discrete. See the caption of Fig. 4 for the correspondence between line types and tests.

We conclude with a mixed case, where the covariates X_1 and X_2 are continuous but the response, Y, is binary. In particular, we let X_1 and X_2 be independent $N(0, 1)$ while $Y \sim \text{Ber}[\text{expit}\{6\sin(X_1X_2)\}]$. Our empirical power computations are displayed in Fig. 6.

No independence test dominates the others in our simulations. The fact that the nonparametric tests often perform nearly as well as, or better than, the distance covariance is surprising, however, as they are invariant with respect to such a wide range of transformations. Indeed, $d_{\text{cov}}^{\text{rank}}$ never performs substantially worse than d_{cov} in our experiments. While it is certainly not a proof, the fact that the tests based on τ_P^* have power beyond the nominal level in all cases suggests that, unlike τ_J^*, perhaps τ_P^* is indeed D-consistent.

Although the use of orthogonal range query data structures reduces the asymptotic complexity of computing our U-statistics of interest, such results give little guidance on which algorithms should be used for realistic sample sizes. In empirical experiments, see the Supplementary Material, we find that U_D, U_R and $U_{\tau_J^*}$ all substantially benefit from the use of a non-naïve approach in practical samples. However, the U-statistic $U_{\tau_P^*}$, probably because of the many large constant factors that are hidden in the asymptotic analysis, is more quickly computed using a naïve approach.

Acknowledgement

This material is based upon work supported by the U.S. National Science Foundation.

Supplementary material

Supplementary material available at *Biometrika* online includes all the proofs, a review of the asymptotic theory of U-statistics, a description of a class of D-consistent summed symmetric rank covariances generalizing Hoeffding's D, and experiments demonstrating the empirical efficiency of our algorithms.

References

Bergsma, W. & Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20**, 1006–28.

Böttcher, B., Keller-Ressel, M. & Schilling, R. L. (2017). Detecting independence of random vectors I. Generalized distance covariance and Gaussian covariance. *arXiv:* 1711.07778.

Breiman, L. & Friedman, J. (1985). Estimating optimal transformations for multiple regression and correlation: Rejoinder. *J. Am. Statist. Assoc.* **80**, 580–98.

Christensen, D. (2005). Fast algorithms for the calculation of Kendall's τ. *Comp. Statist.* **20**, 51–62.

Cui, H., Li, R. & Zhong, W. (2015). Model-free feature screening for ultrahigh dimensional discriminant analysis. *J. Am. Statist. Assoc.* **110**, 630–41.

de Berg, M., Cheong, O., van Kreveld, M. & Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Berlin: Springer, 3rd ed.

Duquenne, V. & Cherfouh, A. (1994). On permutation lattices. *Math. Social Sci.* **27**, 73–89.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O. & Schölkopf, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6**, 2075–129.

Heller, R., Heller, Y. & Gorfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100**, 503–10.

Heller, R., Heller, Y., Kaufman, S., Brill, B. & Gorfine, M. (2016). Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17**, 1–54.

Heller, Y. & Heller, R. (2016). Computing the Bergsma Dassios sign-covariance. *arXiv:* 1605.08732.

Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19**, 546–57.

Hušková, M. & Meintanis, S. G. (2008). Testing procedures based on the empirical characteristic functions I: Goodness-of-fit, testing for symmetry and independence. *Tatra Mt. Math. Publ.* **39**, 225–33.

Kankainen, A. & Ushakov, N. G. (1998). A consistent modification of a test for independence based on the empirical characteristic function. *J. Math. Sci.* **89**, 1486–94.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* **30**, 81–93.

Kinney, J. B. & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Nat. Acad. Sci.* **111**, 3354–9.

Kraskov, A., Stögbauer, H. & Grassberger, P. (2004). Estimating mutual information. *Phys. Rev.* E **69**, article no. 066138.

Leung, D. & Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46**, 280–307.

Ma, L. & Mao, J. (2017). Fisher exact scanning for dependency. *arXiv:* 1608.07885.

Nandy, P., Weihs, L. & Drton, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Statist.* **10**, 2287–311.

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. & Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science* **334**, 1518–24.

Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C. & Mitzenmacher, M. (2016). Measuring Dependence Powerfully and Equitably. *J Mach. Learn. Res*. **17**, 1–63.

Rizzo, M. L. & Szekely, G. J. (2016). *Energy: E-Statistics: Multivariate Inference via the Energy of Data*. R package version 1.7-0, https://CRAN.R-project.org/package=energy.

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101.

Székely, G. J., Rizzo, M. L. & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–94.

Wang, X., Jiang, B. & Liu, J. S. (2017). Generalized R-squared for detecting dependence. *Biometrika* **104**, 129–39.

Weihs, L., Drton, M. & Leung, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Comp. Statist.* **31**, 315–28.

Zhu, L., Xu, K., Li, R. & Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika* **104**, 829–43.

[*Received on* 17 *August* 2017. *Editorial decision on* 7 *February* 2018]