Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Multivariate Rank-Based Distribution-Free Nonparametric Testing Using Measure Transportation

Nabarun Deb & Bodhisattva Sen

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 541, 192–207: Theory and Methods
https://doi.org/10.1080/01621459.2021.1923508

Taylor & Francis
Taylor & Francis Group

Multivariate Rank-Based Distribution-Free Nonparametric Testing Using Measure Transportation

Nabarun Deb and Bodhisattva Sen

Department of Statistics, Columbia University, New York, NY

ABSTRACT

In this article, we propose a general framework for distribution-free nonparametric testing in multi-dimensions, based on a notion of multivariate ranks defined using the theory of measure transportation. Unlike other existing proposals in the literature, these multivariate ranks share a number of useful properties with the usual one-dimensional ranks; most importantly, these ranks are distribution-free. This crucial observation allows us to design nonparametric tests that are exactly distribution-free under the null hypothesis. We demonstrate the applicability of this approach by constructing exact distribution-free tests for two classical nonparametric problems: (I) testing for mutual independence between random vectors, and (II) testing for the equality of multivariate distributions. In particular, we propose (multivariate) rank versions of distance covariance and energy statistic for testing scenarios (I) and (II), respectively. In both these problems, we derive the asymptotic null distribution of the proposed test statistics. We further show that our tests are consistent against all fixed alternatives. Moreover, the proposed tests are computationally feasible and are well-defined under minimal assumptions on the underlying distributions (e.g., they do not need any moment assumptions). We also demonstrate the efficacy of these procedures via extensive simulations. In the process of analyzing the theoretical properties of our procedures, we end up proving some new results in the theory of measure transportation and in the limit theory of permutation statistics using Stein's method for exchangeable pairs, which may be of independent interest.

ARTICLE HISTORY
Received September 2019
Accepted April 2021

KEYWORDS
Asymptotic null distribution; Consistency against fixed alternatives; Distance covariance; Distribution-free inference; Energy distance; Multivariate ranks; Multivariate two-sample testing; Quasi-Monte Carlo sequences; Stein's method for exchangeable pairs; Testing for mutual independence

1. Introduction

Consider the following two classical multivariate nonparametric hypothesis testing problems:

(*I*) *Testing for mutual independence*: Given independent observations from a distribution G on $\mathbb{R}^d$, $d = d_1 + d_2$, $d_1, d_2 \geq 1$, let G_1 and G_2 denote the marginals of G corresponding to the first d_1 and last d_2 components, respectively. Then, the problem of *mutual independence* testing reduces to $\mathrm{H}_0 : G = G_1 \otimes G_2$ versus $\mathrm{H}_1 : G \neq G_1 \otimes G_2$, where by $G_1 \otimes G_2$, we mean the product of the marginal distributions G_1 and G_2. A natural extension of this problem is to test for the mutual independence of K marginals, with $K \geq 2$. The independence testing problem has found applications in a wide variety of disciplines such as in statistical genetics (Liu et al. 2010), survival analysis (Martin and Betensky 2005), ecological risk assessment (Dishion, Capaldi, and Yoerger 1999), independent component analysis (Lu, Lee, and Chiu 2009), etc., and has consequently inspired a long line of research over the past century (see, e.g., Puri and Sen 1971; Gieser and Randles 1997; Hollander, Wolfe, and Chicken 2014, chaps.1 and 8 and the references therein).

(*II*) *Testing for equality of distributions*: Given independent observations from two multivariate distributions, say F_1 and F_2 on $\mathbb{R}^d$, $d \geq 1$, the nonparametric *two-sample goodness-of-fit* testing problem can be formulated as $\mathrm{H}_0 : F_1 = F_2$ versus $\mathrm{H}_1 : F_1 \neq F_2$. The above problem can also be extended to the K-*sample* setup ($K \geq 2$) when one observes independent samples from K distributions and the goal is to nonparametrically test the equality of all the K distributions. The two-sample (or K-sample) problem also has numerous applications, for example, in pharmaceutical studies (Farris and Schopflocher 1999), causal inference (Folkes, Koletsky, and Graham 1987), remote sensing (Conradsen et al. 2003), econometrics (Mayer 1975), etc., and has been studied extensively (see, e.g., Weiss 1960; Bickel 1968; Hollander, Wolfe, and Chicken 2014 and the references therein).

In this article, we study the above two problems and develop *nonparametric* testing procedures that are *exactly distribution-free* (i.e., the null distributions of the test statistics are free of the underlying (unknown) data-generating distributions, for all sample sizes), *computationally feasible* and are *consistent* against all fixed alternatives (i.e., the probability of rejecting the null, calculated under the alternative, converges to 1 as the sample size increases). In fact, we develop a general framework for multivariate distribution-free nonparametric testing applicable much beyond the above two examples. To the best of our knowledge, the test proposed here in the context of testing independence is the first and only nonparametric test that guarantees the three aforementioned desirable properties. In the multivariate two-sample setting, the only two tests with

CONTACT Bodhisattva Sen *bodhi@stat.columbia.edu* Department of Statistics, Columbia University, 1255 Amsterdam Avenue, Room # 1032, New York, NY 10027.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

the above properties are due to Rosenbaum Rosenbaum (2005) and Boeckel, Spokoiny, and Suvorikova (2018).

To construct our finite sample distribution-free tests, we use a suitable notion of *multivariate ranks* (obtained from the theory of measure transportation, to be discussed below) which are themselves distribution-free. This is analogous to what is usually done in one-dimensional problems. Let us illustrate this principle in the context of testing for mutual independence (problem (I)). When $d_1 = d_2 = 1$, the classical product-moment correlation—which mainly captures linear dependence between the variables—can be used to test this hypothesis. However, the exact distribution of the Pearson correlation coefficient, under H_0, depends on the marginals G_1 and G_2. This gave way to Spearman's rank-correlation (another related measure is Kendall's τ coefficient; also see Pearson 1920; Kendall and Gibbons 1990; Gibbons and Chakraborti 2011) which calculates the product-moment correlation between the one-dimensional ranks of the variables. Consequently, the resulting test is distribution-free under the null hypothesis of mutual independence and can deal with nonlinear (monotone) dependencies. Note that the use of ranks to obtain distribution-free tests is ubiquitous in one-dimensional problems in nonparametric statistics—for example, two-sample Kolmogorov–Smirnov test (Smirnoff 1939), Wilcoxon signed-rank test (Wilcoxon 1947), Wald–Wolfowitz runs test (Wald and Wolfowitz 1940), Mann-Whitney rank-sum test (Mann and Whitney 1947), Kruskal–Wallis test (Kruskal 1952), Hoeffding's D-test (Hoeffding 1948), etc.

In the d-dimensional Euclidean space, for $d \geq 2$, due to the absence of a canonical ordering, the existing extensions of concepts like ranks (such as component-wise ranks, e.g., Bickel 1965; Puri and Sen 1966; spatial ranks, e.g., Chaudhuri 1996; Marden 1999; depth-based ranks, e.g., Liu and Singh 1993; Zuo and Serfling 2000; Mahalanobis ranks, e.g., Hallin and Paindaveine 2004, 2006) and the corresponding rank-based tests no longer possess exact distribution-freeness. This raises a fundamental question: "How do we define multivariate ranks that can lead to distribution-free testing procedures?". A major breakthrough in this regard was very recently made in the pioneering work of Marc Hallin and coauthors (Hallin 2017; Hallin et al. 2021; Chernozhukov et al. 2017) where they propose a notion of multivariate ranks, based on the theory of measure transportation, that possesses many of the desirable properties present in their one-dimensional counterparts.

To motivate this notion of multivariate ranks, let us start with the following interpretation of the one-dimensional ranks. Given a collection of n iid random variables $X_1, \ldots, X_n$ on $\mathbb{R}$ (having a continuous distribution) the *rank map* assigns these observations to elements of the set $\{1/n, 2/n, \ldots, n/n\}$ (or $1, 2, \ldots, n$, depending on interpretation) by solving the following optimization problem:

$$\hat{\sigma} := \operatorname{argmin}_{\sigma=(\sigma(1),\ldots,\sigma(n)) \in S_n} \sum_{i=1}^{n} \left| X_i - \frac{\sigma(i)}{n} \right|^2 \qquad (1)$$
$$= \operatorname{argmax}_{\sigma=(\sigma(1),\ldots,\sigma(n)) \in S_n} \sum_{i=1}^{n} \sigma(i) X_i,$$

where S_n is the set of all permutations of $\{1, 2, \ldots, n\}$ (see Villani 2003, chap. 1). It is not difficult to check (by using the rearrangement inequality; see, e.g., Hardy, Littlewood, and Pólya 1952, theor. 368) that $\hat{\sigma}(i)/n$ (or simply $\hat{\sigma}(i)$) will equal the rank of X_i, for $i = 1, \ldots, n$; see the left panel of Figure 1.

Note that Equation (1) can be readily extended to the multivariate setting where the discrete uniform numbers $\{i/n : 1 \leq i \leq n\}$ are replaced by the set of multivariate rank vectors $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\} \subset [0, 1]^d$ — a sequence of "uniform-like" points in $[0, 1]^d$ (see Section E.2 in the supplement for other choices of reference distributions; also see, e.g., Hallin 2017; Hallin et al. 2021; Chernozhukov et al. 2017; Boeckel, Spokoiny, and Suvorikova 2018). In this article, we consider $\{\mathbf{c}_i : 1 \leq i \leq n\}$ as a quasi-Monte Carlo sequence—in particular, we advocate the use of Halton sequences and employ it in our simulation experiments; other natural choices like the equally spaced d-dimensional lattice are also possible (see Section E.3 in the supplement for a detailed discussion). Specifically, given iid random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on $\mathbb{R}^d$, we consider the following optimization problem:

$$\hat{\sigma} := \operatorname{argmin}_{\sigma=(\sigma(1),\ldots,\sigma(n)) \in S_n} \sum_{i=1}^{n} \|\mathbf{X}_i - \mathbf{c}_{\sigma(i)}\|^2 \qquad (2)$$

where, as before, the optimization is over S_n, the set of all permutations of $\{1, 2, \ldots, n\}$, and $\| \cdot \|$ denotes the usual Euclidean norm in $\mathbb{R}^d$. Note that (2) can be viewed as an *assignment problem* (see, e.g., Munkres 1957; Bertsekas 1988) for which algorithms with worst case complexity $\mathcal{O}(n^3)$ are available in the literature (see Appendix B in the supplement for a discussion). Based on Equation (2), one can then define the multivariate rank of $\mathbf{X}_i$ as $\mathbf{c}_{\hat{\sigma}(i)}$. This is illustrated in the right panel of Figure 1 where the dashed lines join the data points (in red) with the corresponding rank vectors (indicated by blue crosses).

The above optimization problem (see Equation (2)) indeed results in a distribution-free notion of empirical multivariate ranks as we demonstrate in Proposition 2.2 (also see (Hallin et al. 2021, prop. 1.6.1)). Note that Equation (2) is connected to the theory of optimal measure transportation as we are "transporting" the empirical distribution of the $\mathbf{X}_i$'s to the empirical distribution of $\mathbf{c}_i$'s. We review this literature and build on the work of Hallin et al. (2021) in Sections 2.1 and 2.2.

Having defined a suitable notion of multivariate ranks, the next natural question becomes: "How does one use these multivariate ranks for nonparametric testing?". In this regard we have a general yet powerful recipe: Given a set of multivariate observations for a nonparametric testing problem (e.g., (*I*) or (*II*)), define their multivariate ranks in such a way (depending on the problem) so that the distribution of these ranks is exactly universal (free of the data generating distribution(s)) under H_0. Next, take a "good" test statistic for the corresponding nonparametric testing problem (which may not be distribution-free under H_0). Then, form a new test by evaluating the original test statistic on these obtained multivariate ranks instead of the data points themselves. Clearly, this will result in a distribution-free test statistic. We believe that this approach is quite general and can consequently be used in a variety of multivariate nonparametric inference problems, much beyond the two problems (*I*) and (*II*) discussed above. As we have observed before, this prescription indeed yields Spearman's rank correlation coefficient when

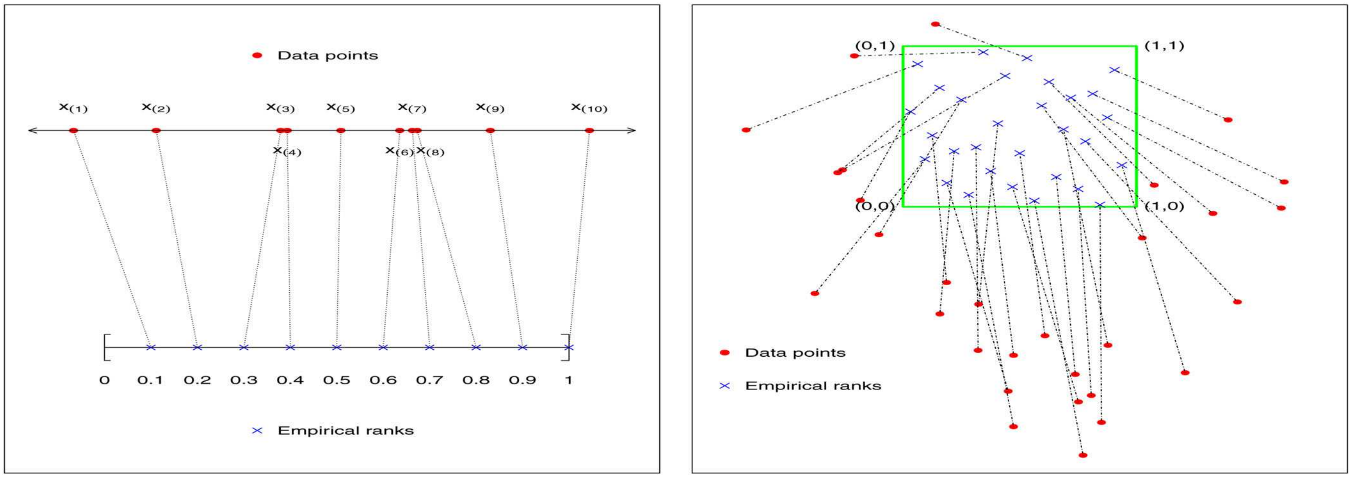

Figure 1. The left panel illustrates the correspondence between univariate data points and their ranks (which are the points i/n, for $i = 1, \ldots, n = 10$). The right panel shows the similar correspondence between bivariate data points and their bivariate ranks which are now pseudo-random numbers in the unit square $[0, 1]^2$. The rank of a data point (in solid red) is given by the blue cross at the other end of the dashed line joining them. Note that the points near the center of the data distribution are mapped close to $(1/2, 1/2)$ whereas the points closer to the extremes of the data cloud are mapped to the corresponding extreme regions of the unit square, thereby giving rise to a natural bivariate ordering of the data points.

applied to the usual product-moment correlation for testing mutual independence when $d_1 = d_2 = 1$. In fact, this general principle has motivated a number of other interesting articles in the last year-and-a-half after the first `arxiv` version of this article was posted (see Section 5, i.e., the discussion section for details).

Let us now describe our contributions in the problem of testing for mutual independence (i.e., problem (*I*)). Over the last 2–3 decades a plethora of nonparametric testing procedures have been proposed for this problem in the multivariate setting; see, for example, Székely, Rizzo, and Bakirov (2007), Berrett and Samworth (2019), Gretton, Fukumizu et al. (2008), Heller, Heller, and Gorfine (2013), Biswas, Sarkar, and Ghosh (2016), Gieser and Randles (1997), Taskinen, Oja, and Randles (2005), Oja (2010), and Friedman and Rafsky (1983) and the references therein. One particular testing procedure, namely *distance covariance* (introduced in Székely, Rizzo, and Bakirov 2007; also see Bakirov, Rizzo, and Székely 2006), has received much attention recently, mainly due to its simplicity and good power properties. We also introduce the "population" version of *rank distance covariance* in Section 3.1 and study its empirical analogue in Section 4.1. In Lemma 3.1, we show that rank distance covariance also characterizes independence.

In Lemma 4.1, we show that our proposed rank distance covariance test is exactly distribution-free as soon as the two marginal distributions are absolutely continuous. In fact, when $d_1 = d_2 = 1$, we show in Lemma 4.2 that our proposed test is exactly equivalent to the statistic put forward in Blum, Kiefer, and Rosenblatt (1961) which in turn is a modification of the celebrated Hoeffding's *D*-statistic (Hoeffding (1948))—one of the first nonparametric tests for mutual independence. We further demonstrate, in Theorem 4.2, that our proposed test is consistent (i.e., has asymptotic power 1) as soon as the two marginals are absolutely continuous. In fact, we do not even need the underlying distributions to have finite means for this result (cf. with usual distance covariance). We also go a step further and obtain the asymptotic distributional limit of our test statistic, under H_0, in Theorem 4.1. This result further demonstrates that the asymptotic limit of our test statistic does not depend on the underlying data-generating distribution and is invariant to the choice of the sequence $\{\mathbf{c}_n\}_{n\geq 1}$—the multivariate ranks.

In Section 4.2, we study the problem of testing for the equality of two multivariate distributions (i.e., problem (*II*)) and propose a test for this goodness of fit using the *rank energy statistic* which is based on the usual energy statistic (as introduced in Székely and Rizzo 2013, also see Baringhaus and Franz 2004; Székely and Rizzo 2005 for definitions and motivation). Similar to distance covariance, the energy statistic is also based on pairwise distances and is easy to compute. The energy statistic equals 0 if and only if the two underlying distributions are the same, as long as the two distributions have finite means. The energy test has also attracted a lot of attention recently in a variety of applications, see, for example, in robust statistics (Klebanov 2002), microarray data analysis (Xiao et al. 2004), material structure analysis (Beneš et al. 2009), etc.

We demonstrate the distribution-free nature of our proposed test statistic for problem (*II*) in Lemma 4.3. An interesting property of this proposed statistic is that it is exactly equivalent to the famous two-sample Cramér-von Mises statistic (see, e.g., Anderson 1962 when $d = 1$. We explain this connection in Lemma 4.4. We further prove the consistency and derive the asymptotic distribution (under H_0) of our proposed rank-based energy test statistic in Theorems 4.4 and 4.3, respectively. The population version of this rank-based energy statistic exhibits several interesting and desirable properties which we highlight in Lemma 3.2.

Moreover, we extend both the above tests to their multi-sample versions in Appendix E.1 in the supplement; the corresponding theoretical results are presented in Propositions E.1 and E.2.

In Appendix C, we carry out extensive simulation experiments to study the power behavior of the proposed tests. These simulations show that our proposed procedures for independence testing and two-sample goodness-of-fit testing perform well under a variety of alternatives, often outperforming competing methods. In general, these distribution-free tests have good efficiency, are more powerful for distributions with heavy tails and are more robust to outliers and contaminations. In Appendix A (in the supplement), we demonstrate practical advantages of our proposals over competing methods via the analysis of two benchmark datasets.

In the following, we encapsulate some of the main contributions of this article, all the while comparing our procedures to existing approaches from the statistics literature.

(i) *Exact distribution-freeness*: As mentioned before, our proposals are all exactly distribution-free in finite samples. This is a particularly desirable property as it avoids the need to estimate any nuisance parameters, or use resampling/permutation ideas, or conservative asymptotic approximations, for determining rejection thresholds. Moreover, distribution-free procedures can help reduce computational burden in statistical problems—a very practical concern in this era of big data; see, e.g., (Heller, Gorfine, and Heller 2012, sec. 7) for an interesting discussion on this topic. As far as we are aware, the only distribution-free methods available in the literature for tackling the above discussed problems (*I*) and (*II*) are: Rosenbaum (2005),Boeckel, Spokoiny, and Suvorikova (2018), Biswas, Mukhopadhyay, and Ghosh (2014) for the multivariate two-sample problem and Biswas, Sarkar, and Ghosh (2016), Heller, Gorfine, and Heller (2012), and Heller and Heller (2016) for the mutual independence testing problem. Note that exact distribution-freeness is stronger than obtaining distribution-free thresholds for testing procedures (which are usually conservative) as in Gretton and Györfi (2008), Gretton and Györfi (2010), Gretton et al. (2012), and Biau and Györfi (2007).

(ii) *Completely nonparametric and computationally feasible*: Being based on multivariate ranks, our proposal is completely nonparametric. Moreover, our proposed test statistics can be computed with worst-case complexity $\mathcal{O}(n^3)$ for all dimensions (once the pairwise distances between data points are calculated). Further, our procedures do not depend crucially on the choice of the $\mathbf{c}_i$'s (as in Equation (2)); see Theorems 4.1 and 4.3. In Appendix B (in the supplement), we explain how our proposed test statistics can be computed in a few simple steps using readily available R packages. Although exactly distribution-free graph-based tests for mutual independence and two-sample goodness-of-fit testing were proposed in Biswas, Sarkar, and Ghosh (2016) and Biswas, Mukhopadhyay, and Ghosh (2014), respectively, these tests are extremely expensive to compute and possibly not applicable even for moderate sample sizes.

(iii) *Consistency under absolute continuity*: The only condition we need on the underlying distributions for the consistency of our tests is that they are absolutely continuous (no moment conditions are necessary). This enables their direct usage for nonparametric inference under heavy-tailed data-generating distributions such as stable laws Yang (2012) and Pareto distributions Rizzo (2009), and also sets them apart from popular methods such as usual distance covariance and energy statistic. To the best of our knowledge, there are only two computationally efficient exactly distribution-free multivariate mutual independence testing procedures in literature, both based on a similar graph-based framework and proposed simultaneously in Heller, Gorfine, and Heller (2012); also see Heller and Heller (2016). However, in Heller, Gorfine, and Heller (2012), the authors did not provide any results that guarantee consistency of their tests against fixed alternatives.

(iv) *Broader scope of applications in multivariate nonparametric testing*: As described before, our approach is holistic. Based on our ideas, one can easily construct multivariate rank-based distribution-free tests for mutual independence using other statistics, such as Hilbert-Schmidt independence criteria (Gretton, Fukumizu et al. 2008) or HHG (Heller, Heller, and Gorfine 2013), instead of distance covariance; same goes for the goodness-of-fit testing problem. Note that, although we delve deep into these two particular nonparametric problems, we essentially describe a general principle to construct distribution-free tests in multivariate nonparametric settings that can be used in a variety of other contexts; e.g., in tests of symmetry (Székely and Rizzo 2013), hierarchical clustering (Székely and Rizzo 2005), change point analysis (Székely and Rizzo 2009), etc.

The rest of the article is organized as follows. In Section 2, we start with a brief overview of measure transportation (Section 2.1), followed by a description of our proposed multivariate ranks and their properties (Sections 2.2 and 2.3). Section 3 introduces new measures of multivariate association and goodness of fit and also discusses some interesting properties of these measures that make them desirable. Our proposed procedures for testing mutual independence and equality of distributions are introduced in Section 4 (along with their multi-sample extensions). In that section, we also discuss interesting/useful properties of our test statistics and provide theoretical guarantees with regards to distribution-freeness, consistency and asymptotic null distribution. We conclude the main article with a brief discussion in Section 5. Appendices C, D and A (in the supplement) illustrate the usefulness of our proposed methods via simulation experiments and real data analysis. We conclude the main article with a brief discussion in Section 5. In Appendix B (in the supplement), we explain how the proposed test statistics can be computed using standard software packages (in R). Appendix E.3 (in the supplement) is aimed at providing a very brief introduction to the field of quasi-Monte Carlo methods which plays a tangential role in our approach. Finally, in Appendices F and G (see the supplement) we provide the proofs of our main results, while in Appendix H (see the supplement), we discuss some existing results on convex analysis and Stein's method of exchangeable pairs, which are used in the proofs of our main results.

The methods described in the article have been implemented using the R software. The relevant codes, including simulation experiments, are available in the first author's GitHub page.

After the first version of this article was posted on arxiv, we were made aware of the article (Shi, Drton, and Han 2020) (uploaded a few days after our first submission on arxiv). Shi, Drton, and Han (2020) considered distribution-free mutual independence testing of two random vectors (i.e., problem (I)) using multivariate ranks as described in Hallin et al. (2021). The article also shows the distribution-freeness and consistency of the same test-statistic as in Section 4.1 of this article. However, the asymptotic consistency results in Shi, Drton, and Han (2020) are derived under more stringent conditions (e.g., nonvanishing Lebesgue probability densities). Note that in our article, we develop a general framework for multivariate distribution-free nonparametric testing using optimal transportation, applicable much beyond problem (I); in particular, we also consider problem (II).

2. Multivariate Ranks and Quantiles

In this section, we define ranks and quantiles for multivariate distributions (both population and empirical versions) using the theory of measure transportation; our approach is similar to that of Hallin et al. (2021) and Boeckel, Spokoiny, and Suvorikova (2018). This will serve a pivotal role in defining the test statistics that appear later in the article.

2.1. Preliminaries: Overview of Measure Transportation

Let us introduce some notation for the rest of the article. We will use $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ to denote the standard Euclidean norm and inner product on a suitable finite-dimensional Euclidean space (say $\mathbb{R}^d$), respectively. Weak convergence of distributions will be denoted by $\stackrel{w}{\to}$ while $\stackrel{d}{=}$ will denote equality in distribution. We will use $\mathcal{U}^d$ to denote the uniform distribution on $[0,1]^d$, and S_n for the set of all permutations of $\{1,2,\ldots,n\}$. Let $\delta_{\mathbf{a}}$ denote the Dirac measure that assigns probability 1 to the point $\mathbf{a}$. Finally, let $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_{ac}(\mathbb{R}^d)$ denote the families of all probability distributions and Lebesgue absolutely continuous probability measures on $\mathbb{R}^d$, respectively.

As the name suggests, "measure transportation" (perhaps more commonly referred to as *optimal transportation*) is the problem of finding "nice" functions $\mathbf{F}:\mathbb{R}^d\to\mathbb{R}^d$ such that $\mathbf{F}$ pushes a given measure $\mu\in\mathcal{P}(\mathbb{R}^d)$ to $\nu\in\mathcal{P}(\mathbb{R}^d)$. Here, by $\mathbf{F}$ pushes μ to ν, usually written as $\mathbf{F}\#\mu=\nu$, we mean that $\mathbf{F}(\mathbf{X})\sim\nu$ where $\mathbf{X}\sim\mu$. This rich area of mathematics was initiated by the work of Gaspard Monge in 1781 (see Monge 1781). Based on already introduced notation, perhaps the simplest version of *Monge's problem* is as follows:

$$\inf_{\mathbf{F}}\int\|\mathbf{x}-\mathbf{F}(\mathbf{x})\|^2\,d\mu(\mathbf{x})\qquad\text{subject to}\quad \mathbf{F}\#\mu=\nu; \tag{3}$$

this is technically a mis-characterization as Monge originally worked with the loss $\|\cdot\|$ instead of $\|\cdot\|^2$. A minimizer of (3), if it exists, is referred to as an *optimal transport map*. One of the most powerful results in this field came into being from Brenier's *polar factorization theorem* (see Brenier 1991) which yields: If $\mu,\nu\in\mathcal{P}_{ac}(\mathbb{R}^d)$ have finite second-order moments, then the corresponding Monge's problem admits a μ-a.e. unique solution which happens to be the gradient of a convex function.

While the above approach addresses the problem of finding functions that push μ to ν, the assumption on the second-order moments (which is a basic requirement for Monge's problem to make sense) seems extraneous and inappropriate. Indeed, for $d=1$, if F_μ and F_ν are the distribution functions associated with μ and ν (assumed to be absolutely continuous), respectively, then $F_\nu^{-1}\circ F_\mu$ pushes μ to ν without any moment assumptions. A ground-breaking extension of this univariate property was proved by McCann (1995), where he took a geometric approach to the problem of measure transportation. His result is the defining tool we will need to make sense of the definitions in this section. Therefore, let us state *McCann's theorem* in a form which will be useful to us; see, for example, Villani (2003, theor. 2.12 and Corol 2.30).

Proposition 2.1 (McCann's theorem (McCann 1995)). Suppose that $\mu,\nu\in\mathcal{P}_{ac}(\mathbb{R}^d)$. Then, there exist functions $\mathbf{R}(\cdot)$ and $\mathbf{Q}(\cdot)$ (hereafter referred to as "transport maps"), both of which are gradients of (extended) real-valued d-variate convex functions (hereafter called "transport potentials"), such that $\mathbf{R}\#\mu=\nu$, $\mathbf{Q}\#\nu=\mu$, $\mathbf{R}$ and $\mathbf{Q}$ are unique (μ and ν a.e., respectively), $\mathbf{R}\circ\mathbf{Q}(\mathbf{x})=\mathbf{x}$ (μ a.e.) and $\mathbf{Q}\circ\mathbf{R}(\mathbf{y})=\mathbf{y}$ (ν a.e.). Moreover, if μ and ν have finite second moments, $R(\cdot)$ is also the solution to Monge's problem in Equation (3).

Observe that McCann's theorem does away with all moment assumptions and guarantees existence and (a.e.) uniqueness of *transport maps* under minimal assumptions on μ and ν. Note that any convex function on $\mathbb{R}^d$ is differentiable Lebesgue a.e., and consequently μ (or ν) a.e. by Alexandroff theorem (see, e.g., Alexandroff 1939). In Proposition 2.1, by "gradient of a convex function," we essentially refer to a function from $\mathbb{R}^d\to\mathbb{R}^d$ which is μ (or ν) a.e. equal to the gradient of some convex function.

2.2. Definitions of Multivariate Ranks

Definition 2.1 (Population multivariate ranks and quantiles). Set $\nu=\mathcal{U}^d$. Given a measure $\mu\in\mathcal{P}_{ac}(\mathbb{R}^d)$, the corresponding population rank and quantile maps are defined as functions $\mathbf{R}(\cdot)$ and $\mathbf{Q}(\cdot)$, respectively (as in Proposition 2.1). Note that these are unique only up to measure zero sets with respect to μ and ν, respectively.

Remark 2.1. The smoothness and regularity properties of the population rank and quantile maps as in Definition 2.1 have been studied extensively over the past 30 years or so. Since such discussions are beyond the scope of this article, we would like to refer the interested reader to De Philippis and Figalli (2013), Caffarelli (1990), and Villani (2009, chap. 12).

In standard statistical applications, the population rank map is not available to the practitioner. In fact, the only accessible information about the measure μ comes in the form of empirical observations $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n\stackrel{\text{iid}}{\sim}\mu\in\mathcal{P}_{ac}(\mathbb{R}^d)$. A natural question thus arises: "How can we estimate population ranks from empirical observations?". In this direction, let

$$\mathcal{H}_n^d:=\{\mathbf{h}_1^d,\ldots,\mathbf{h}_n^d\} \tag{4}$$

denote the (fixed) set of sample *multivariate rank vectors* (analogous to $\mathbf{c}_i$'s in (2)). In practice, for $d\geq 2$ we may take $\mathcal{H}_n^d$ to be the d-dimensional Halton sequence of size n (or any quasi-Monte Carlo sequence; see Appendix E.3 in the supplement for more details), and the usual $\{i/n\}_{1\leq i\leq n}$ sequence when $d=1$. The empirical distribution on $\mathcal{H}_n^d$ will serve as a discrete approximation of $\mathcal{U}^d$. Also, let $\mathcal{D}_n^X:=\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$ be the observed data. Let

$$\mu_n^{\mathbf{X}}:=\frac{1}{n}\sum_{i=1}^n\delta_{\mathbf{X}_i}\qquad\text{and}\qquad \nu_n:=\frac{1}{n}\sum_{i=1}^n\delta_{\mathbf{h}_i^d} \tag{5}$$

denote the empirical distributions on $\mathcal{D}_n^X$ and $\mathcal{H}_n^d$, respectively.

Definition 2.2 (Empirical rank map). We define the empirical rank function $\hat{\mathbf{R}}_n : \mathcal{D}_n^X \to \mathcal{H}_n^d$ as the optimal transport map which transports $\mu_n^{\mathbf{X}}$ (the empirical distribution on the data) to ν_n (the empirical distribution on $\mathcal{H}_n^d$), that is,

$$\hat{\mathbf{R}}_n = \text{argmin}_{\mathbf{F}} \int \|\mathbf{x} - \mathbf{F}(\mathbf{x})\|^2 \, d\mu_n^{\mathbf{X}}(\mathbf{x}) \qquad \text{subject to}$$
$$\mathbf{F}\#\mu_n^{\mathbf{X}} = \nu_n \tag{6}$$

Note that Equation (6) can be thought of as the discrete analog of Equation (3) which defines the population rank function $\mathbf{R}(\cdot)$ if μ has finite second moments. Further, Equation (6) is equivalent to the following optimization problem:

$$\hat{\sigma}_n := \text{argmin}_{\sigma \in S_n} \sum_{i=1}^{n} \|\mathbf{X}_i - \mathbf{h}_{\sigma(i)}^d\|^2 = \text{argmax}_{\sigma \in S_n} \sum_{i=1}^{n} \langle \mathbf{X}_i, \mathbf{h}_{\sigma(i)}^d \rangle. \tag{7}$$

The equivalence between the two optimization problems in Equation (7) can be easily established by writing out the norms in terms of standard inner products. Note that $\hat{\sigma}_n$ is a.s. uniquely defined (for each n). Now, based on Equation (7), observe that the sample rank map $\hat{\mathbf{R}}_n$ satisfies

$$\hat{\mathbf{R}}_n(\mathbf{X}_i) = \mathbf{h}_{\hat{\sigma}_n(i)}^d, \qquad \text{for } i = 1, \ldots, n. \tag{8}$$

Remark 2.2. The optimization problem in Equation (7) is a combinatorial optimization problem. However, it is known to be equivalent to a linear program and can consequently be solved by standard solvers. Moreover, the special structure of the above problem allows us to view it as an assignment problem (see Munkres 1957; Bertsekas 1988) for which algorithms with worst case complexity $\mathcal{O}(n^3)$ are available in the literature. We will discuss this in more detail in Appendix B in the supplement.

Remark 2.3 (Connection to usual ranks in one-dimension). For $d = 1$, if we use $\mathcal{H}_n^1 = \{i/n\}_{i=1}^n$, then the empirical ranks $\hat{\mathbf{R}}_n(\cdot)$ reduce to the usual notion of one-dimensional ranks.

Remark 2.4 (Why choose a quasi-Monte Carlo sequence?). In this article, we use a quasi-Monte Carlo sequence to compute the multivariate empirical rank map. There are many reasons to prefer a quasi-Monte Carlo sequence as: (i) it is deterministic, (ii) can be constructed for every n and d, (iii) the sequence need not be recomputed if n increases by 1, and importantly, (iv) it provides greater uniformity as it is a low-discrepancy sequence (the discrepancy of a sequence is low if the proportion of points in the sequence falling into an arbitrary hyperrectangle is close to the Lebesgue measure of the set); see Appendix E.3 for more details. In our numerical computations, we use the Halton sequence—a quasi-Monte Carlo sequence—which is readily available in computing packages; for example, in the `R` package `randtoolbox`. Note that the equally-spaced d-dimensional lattice does not satisfy (ii) and (iii) above.

2.3. Properties of Multivariate Ranks

When $d = 1$, the notion of ranks has a number of desirable properties which have been useful in analyzing rank-based estimators and test statistics (see, e.g., Hallin et al. 2021, Part I and the references therein). Below in Proposition 2.2 (proved in Appendix F.1 in the supplement), we reproduce some of these properties for the empirical multivariate ranks as in Definition 2.2. Proposition 2.2 is in fact very similar to (Hallin et al. 2021, prop. 1.6.1) (also see Hallin 2017, prop. 6.1), with some differences which we will elaborate in Appendix E.2 in the supplement.

Proposition 2.2. Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} \mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. We define an *order statistic* $\mathbf{X}_{(\cdot)}^{(n)}$ of $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ as any fixed, arbitrary ordered version of the same—for example, $\mathbf{X}_{(\cdot)}^{(n)} = (\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$, where $\mathbf{X}_{(i)}$ is such that the first coordinate of $\mathbf{X}_{(i)}$ is the ith-order statistic of the n-tuple formed by the first coordinates of the n-vectors in $\mathbf{X}_{(\cdot)}^{(n)}$. Then

(i) The *order statistic* $\mathbf{X}_{(\cdot)}^{(n)}$ is complete and sufficient.
(ii) The vector $(\hat{\mathbf{R}}_n(\mathbf{X}_1), \ldots, \hat{\mathbf{R}}_n(\mathbf{X}_n))$ is uniformly distributed over the $n!$ permutations of the fixed grid $\mathcal{H}_n^d$ (see (4)).
(iii) $(\hat{\mathbf{R}}_n(\mathbf{X}_1), \ldots, \hat{\mathbf{R}}_n(\mathbf{X}_n))$ and $\mathbf{X}_{(\cdot)}^{(n)}$ are mutually independent.

Remark 2.5 (On Proposition 2.2). Property (ii) from Proposition 2.2 is an analog of the distribution-freeness of one-dimensional ranks. Property (iii) may be interpreted as the independence between ranks and order statistics.

As we will see in Section 4, the distribution-free property of the empirical multivariate ranks will lead to the distribution-freeness of the proposed test statistics. However, to guarantee the consistency of the proposed tests, we need the sample rank maps to be well-behaved as the sample size grows. In fact, in the following theorem (proved in Appendix F.2 in the supplement) we show that the sample rank map converges to its population counterpart (i.e., the population rank function $\mathbf{R}(\cdot)$ as in Definition 2.1) in a suitable sense, under minimal assumptions.

Theorem 2.1 (Almost sure L^2 convergence). Assume $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} \mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. Suppose that $\nu_n \xrightarrow{w} \mathcal{U}^d$; see Equation (5). Then $\frac{1}{n}\sum_{i=1}^{n} \|\hat{\mathbf{R}}_n(\mathbf{X}_i) - \mathbf{R}(\mathbf{X}_i)\| \xrightarrow{\text{a.s.}} 0$.

Remark 2.6 (Convergence under L^p-norm). As $\hat{\mathbf{R}}_n(\cdot)$ and $\mathbf{R}(\cdot)$ are uniformly bounded, Theorem 2.1 implies convergence with respect to any L^p-norm, for $1 \le p < \infty$.

Remark 2.7 (On absolute continuity of μ). It is perhaps instructive to note that, even for ranks in one-dimension, the distribution-free property does not hold if the data generating measure is not continuous. Since distribution-free inference is the main goal of this article, it seems reasonable to assume absolute continuity of μ.

It is easy to see that the a.s.-convergence, as presented in Theorem 2.1, is weaker than *uniform convergence*; see, for example, Ghosal and Sen (2019), Chernozhukov et al. (2017), and Hallin et al. (2021). However, the assumption in Theorem 2.1 is minimal—we only assume the absolute continuity of μ, which is much weaker than the assumptions in the above references. A related result that proves a local uniform convergence of empirical rank maps can be found in Zemel and Panaretos (2019, prop.

6) where the authors additionally impose finite second moment assumptions on μ. In Section 4, we will highlight specifically how and why Theorem 2.1 provides a more useful notion of convergence necessary for the results in this article.

3. New Multivariate Rank-Based Measures for Nonparametric Testing

We introduce new multivariate rank-based measures of dependence and goodness of fit in this section and study the properties of these population quantities.

3.1. Rank-Based Dependence Measure

To motivate our proposal, let us start with $d = 1$. Suppose that Z_1 and Z_2 are real-valued absolutely continuous random variables with distribution functions $G_1(\cdot)$ and $G_2(\cdot)$. It is a simple probability exercise to show that Z_1 and Z_2 are independent if and only if $G_1(Z_1)$ and $G_2(Z_2)$ are independent (a more general version of this result will be proved later in the article, see Lemma 3.1 (part (b)). Thus, Z_1 and Z_2 are independent if and only if the joint characteristic function of $(G_1(Z_1), G_2(Z_2))$ factors as the product of the marginal characteristic functions, that is, for all $(t, s) \in \mathbb{R}^2$,

$$\big|\mathbb{E}\exp\big(itG_1(Z_1) + isG_2(Z_2)\big) - \mathbb{E}\exp\big(itG_1(Z_1)\big)\mathbb{E}\exp\big(isG_2(Z_2)\big)\big|^2 = 0.$$

This suggests the following natural measure of dependence:

$$\mathcal{R}_w := \int\int \big|\mathbb{E}\exp\big(iG_{t,s}(\mathbf{Z})\big) - \mathbb{E}\exp\big(itG_1(Z_1)\big)\mathbb{E}\exp\big(isG_2(Z_2)\big)\big|^2 w(t,s)\, dt\, ds,$$

where $\mathbf{Z} = (Z_1, Z_2)$, $G_{t,s}(\mathbf{Z}) = tG_1(Z_1) + sG_2(Z_2)$ and $w : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is a weight function such that $\mathcal{R}_w < +\infty$. The following proposition (proved in Appendix F.3 in the supplement) draws a connection between $\mathcal{R}_w$ and the classical Spearman's rank correlation, which we think has not been observed before.

Proposition 3.1. Consider the notation introduced above. Set $f_{Z_1,Z_2}(t,s) := \mathbb{E}\exp\big(itG_1(Z_1) + isG_2(Z_2)\big) - \mathbb{E}\exp\big(itG_1(Z_1)\big)\mathbb{E}\exp\big(isG_2(Z_2)\big)$. Then,

$$\lim_{t,s\to 0, |t|/|s|\to c} \frac{\big|f_{Z_1,Z_2}(t,s)\big|^2}{\big|f_{Z_1,Z_1}(t,s)\big|\big|f_{Z_2,Z_2}(t,s)\big|} = \rho^2\big(G_1(Z_1), G_2(Z_2)\big), \tag{9}$$

where $\rho^2\big(G_1(Z_1), G_2(Z_2)\big)$ denotes the usual correlation between $G_1(Z_1)$ and $G_2(Z_2)$. In the above display, $c > 0$ is finite, and ensures that s and t do not converge to 0 at "different rates."

The right-hand side of Equation (9) may be interpreted as the population analogue of the classical *Spearman's rank correlation*. Note that applications of Spearman's rank correlation as a measure of association have been extensively studied in the statistics literature (see, e.g., Hauke and Kossowski 2011; Iman and Conover 1982; Mukaka 2012).

Remark 3.1. Proposition 3.1 shows how Spearman's rank correlation measure effectively looks at the difference between the joint and marginal characteristic functions for small (in magnitude) choices of t and s. Therefore, $\mathcal{R}_w$ (after rescaling) offers a very natural extension to Spearman's rank correlation, but it can capture all kinds of departures from independence.

Remark 3.2. It is easy to see that the right-hand side of Equation (9) being 0 does not imply the independence of Z_1 and Z_2. For example, say $Z_1 \sim \mathcal{U}^1$ and $Z_2 = Z_1$ if $Z_1 \in [1/4, 3/4]$, $Z_2 = 1 - Z_1$ if $Z_1 \in (0, 1/4) \cup (3/4, 1)$. Then, $Z_2 \sim \mathcal{U}^1$ and both $G_1(\cdot)$, $G_2(\cdot)$ are identity functions on $(0, 1)$. Therefore, $\mathbb{E}[G_1(Z_1)G_2(Z_2)] - \mathbb{E}[G_1(Z_1)]\mathbb{E}[G_2(Z_2)] = \mathbb{E}[Z_1Z_2] - 1/4 = 0$.

The above discussion now raises the following two questions: "Can we extend $\mathcal{R}_w$ beyond $d = 1$? Also, how do we choose the weight function $w(\cdot,\cdot)$?". For the first question, we will proceed by replacing $G_1(\cdot)$ and $G_2(\cdot)$ with the notion of population multivariate ranks as introduced in Definition 2.1. For the second question, we will borrow the weight function from the seminal article Székely, Rizzo, and Bakirov (2007) where the authors introduced the notion of *distance covariance*. As in Székely, Rizzo, and Bakirov (2007), we do not make any claims on the optimality of our proposed weight function except that it ensures simple, applicable empirical formulae and an exact equivalence between $\mathcal{R}_w$ and the independence between Z_1 and Z_2. We are now in a position to formally define the new rank-based multivariate measure of dependence.

Definition 3.1 (Rank distance covariance). Suppose that $\mathbf{Z}_1 \sim \mu_1$ and $\mathbf{Z}_2 \sim \mu_2$ (not necessarily independent) such that $\mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^{d_1})$ and $\mu_2 \in \mathcal{P}_{ac}(\mathbb{R}^{d_2})$. Let $\mathbf{R}_1(\cdot)$ and $\mathbf{R}_2(\cdot)$ denote the corresponding population rank maps (Definition 2.1). The rank distance covariance (RdCov^2) between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is defined as the usual distance covariance between $\mathbf{R}_1(\mathbf{Z}_1)$ and $\mathbf{R}_2(\mathbf{Z}_2)$, that is,

$$\mathrm{RdCov}^2(\mathbf{Z}_1, \mathbf{Z}_2) := \int_{\mathbb{R}^{d_1+d_2}} \times \frac{\big|\mathbb{E}\exp\big(i\mathbf{R}_{\mathbf{t},\mathbf{s}}(\mathbf{Z})\big) - \mathbb{E}\exp\big(i\mathbf{t}^\top\mathbf{R}_1(\mathbf{Z}_1)\big)\mathbb{E}\exp\big(i\mathbf{s}^\top\mathbf{R}_2(\mathbf{Z}_2)\big)\big|^2}{c(d_1)c(d_2)\|\mathbf{t}\|^{1+d_1}\|\mathbf{s}\|^{1+d_2}} \times d\mathbf{t}\, d\mathbf{s}, \tag{10}$$

where $\mathbf{Z} := (\mathbf{Z}_1, \mathbf{Z}_2)$, $\mathbf{R}_{\mathbf{t},\mathbf{s}}(\mathbf{Z}) := \mathbf{t}^\top\mathbf{R}_1(\mathbf{Z}_1) + \mathbf{s}^\top\mathbf{R}_2(\mathbf{Z}_2)$ and $c(d) := \pi^{(1+d)/2}\big(\Gamma((1+d)/2)\big)^{-1}$.

Definition 3.2 (Rank distance correlation). The rank distance correlation (RdCorr) between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is defined as the usual distance correlation (see Székely, Rizzo, and Bakirov 2007, eq. 2.7) between $\mathbf{R}_1(\mathbf{Z}_1)$ and $\mathbf{R}_2(\mathbf{Z}_2)$. In other words,

$$\mathrm{RdCorr}^2(\mathbf{Z}_1, \mathbf{Z}_2) := \frac{\mathrm{RdCov}^2(\mathbf{Z}_1, \mathbf{Z}_2)}{\mathrm{RdCov}(\mathbf{Z}_1, \mathbf{Z}_1)\mathrm{RdCov}(\mathbf{Z}_2, \mathbf{Z}_2)}. \tag{11}$$

$\mathrm{RdCorr}^2(\mathbf{Z}_1, \mathbf{Z}_2)$ is well-defined by Lemma 3.1 (part (c)). By (Székely, Rizzo, and Bakirov 2007, theor. 3), it follows directly that $\mathrm{RdCorr}(\mathbf{Z}_1, \mathbf{Z}_2) \in [0, 1]$.

Now let us look into some of the properties of RdCov that make it a desirable measure of dependence. The proof of the following lemma is given in Appendix F.4 (in the supplement).

Lemma 3.1. Under the same assumptions as in Definition 3.1, we have

(a) Suppose that $(\mathbf{Z}_1^1, \mathbf{Z}_2^1), (\mathbf{Z}_1^2, \mathbf{Z}_2^2), (\mathbf{Z}_1^3, \mathbf{Z}_2^3)$ are independent observations having the same distribution as $(\mathbf{Z}_1, \mathbf{Z}_2)$. Then,

$$\begin{aligned}
&\mathrm{RdCov}^2(\mathbf{Z}_1, \mathbf{Z}_2) \\
&\quad= \mathbb{E}\big[\|\mathbf{R}_1(\mathbf{Z}_1^1) - \mathbf{R}_1(\mathbf{Z}_1^2)\|\|\mathbf{R}_2(\mathbf{Z}_2^1) - \mathbf{R}_2(\mathbf{Z}_2^2)\|\big] \\
&\qquad+ \mathbb{E}\big[\|\mathbf{R}_1(\mathbf{Z}_1^1) - \mathbf{R}_1(\mathbf{Z}_1^2)\|\big]\mathbb{E}\big[\|\mathbf{R}_2(\mathbf{Z}_2^1) - \mathbf{R}_2(\mathbf{Z}_2^2)\|\big] \\
&\qquad- 2\mathbb{E}\big[\|\mathbf{R}_1(\mathbf{Z}_1^1) - \mathbf{R}_1(\mathbf{Z}_1^2)\|\|\mathbf{R}_2(\mathbf{Z}_2^1) - \mathbf{R}_2(\mathbf{Z}_2^3)\|\big].
\end{aligned} \tag{12}$$

(b) $\mathrm{RdCov}(\mathbf{Z}_1, \mathbf{Z}_2) = 0$ if and only if $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent.
(c) $\mathrm{RdCov}(\mathbf{Z}_1, \mathbf{Z}_1) > 0$.
(d) (Invariance) Suppose $\mathbf{a}_1 \in \mathbb{R}^{d_1}$, $\mathbf{a}_2 \in \mathbb{R}^{d_2}$ and $b_1, b_2 > 0$. Then, $\mathrm{RdCorr}(\mathbf{Z}_1, \mathbf{Z}_2) = \mathrm{RdCorr}(\mathbf{a}_1 + b_1\mathbf{Z}_1, \mathbf{a}_2 + b_2\mathbf{Z}_2)$.
(e) Suppose that $(\mathbf{Z}_1^n, \mathbf{Z}_2^n) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is a sequence of random vectors that converge weakly to $(\mathbf{Z}_1, \mathbf{Z}_2)$; here we assume that $\mathbf{Z}_1^n$ and $\mathbf{Z}_2^n$ have absolutely continuous distributions for all n. Then, $\mathrm{RdCov}^2(\mathbf{Z}_1^n, \mathbf{Z}_2^n) \longrightarrow \mathrm{RdCov}^2(\mathbf{Z}_1, \mathbf{Z}_2)$ as $n \to \infty$.

We would like to refer the interested reader to Móri and Székely (2019) for an elaborate discussion on the importance of these properties in a dependence measure.

Remark 3.3. Unlike distance covariance (see Székely, Rizzo, and Bakirov 2007; Móri and Székely 2019), Lemma 3.1 does not require any moment assumptions on $\mathbf{Z}_1$ and $\mathbf{Z}_2$. However, we do need absolute continuity of the underlying measures μ_1 and μ_2, an assumption which has been justified in Remark 2.7.

Remark 3.4. In Székely, Rizzo, and Bakirov (2007, theor. 7), a closed-form expression for distance covariance when (X, Y) has a bivariate normal distribution, parameterized by correlation ρ, is derived. Although for rank distance covariance (as defined in (10)) such a closed-form expression is not easy to obtain, we can readily approximate it using Monte Carlo. In Appendix D.3 (see the supplement), we demonstrate that, in this bivariate normal setting, the population distance covariance and population rank distance covariance are both monotone in $|\rho|$ and essentially indistinguishable as functions of ρ.

3.2. Rank-Based Measure for Two-Sample Goodness of Fit

We can use a similar approach as in Section 3.1 to come up with a measure for multivariate two-sample goodness-of-fit testing. Define $\mathcal{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ and let $\kappa(\cdot)$ denote the uniform measure on $\mathcal{S}^{d-1}$. Further, assume $\mathbf{Z}_1 \sim \mu_1$ and $\mathbf{Z}_2 \sim \mu_2$ are independent where $\mu_1, \mu_2 \in \mathcal{P}_{ac}(\mathbb{R}^d)$. Then, using the continuity and uniqueness of characteristic functions, it is rather straightforward to check that $\mu_1 = \mu_2$ if and only if $\mathbf{a}^\top\mathbf{Z}_1 \overset{d}{=} \mathbf{a}^\top\mathbf{Z}_2$ for κ a.e. $\mathbf{a}$ (for more details see (Baringhaus and Franz 2004, theorem 2.1)). Therefore, a natural way to measure equality of distributions $\mu_1 = \mu_2$ would be to compare $\mathbb{P}(\mathbf{a}^\top\mathbf{Z}_1 \le t)$ and $\mathbb{P}(\mathbf{a}^\top\mathbf{Z}_2 \le t)$ for all $\mathbf{a} \in \mathcal{S}^{d-1}$ and all $t \in \mathbb{R}$. This provides the main motivation behind the *energy measure* for two-sample goodness of fit (see Baringhaus and Franz 2004; Székely and Rizzo 2013), which is defined as follows:

$$\mathrm{En}(\mathbf{Z}_1, \mathbf{Z}_2) := \gamma_d \int_{\mathbb{R}} \int_{\mathcal{S}^{d-1}} \Big(\mathbb{P}(\mathbf{a}^\top\mathbf{Z}_1 \le t) - \mathbb{P}(\mathbf{a}^\top\mathbf{Z}_2 \le t)\Big)^2 \, d\kappa(\mathbf{a})\, dt$$

where $\gamma_d := \big(2\Gamma(d/2)\big)^{-1}\sqrt{\pi}(d-1)\Gamma\big((d-1)/2\big)$ for $d > 1$ and $\gamma_d := 1$ for $d = 1$. It can be shown that $\mathrm{En}(\mathbf{Z}_1, \mathbf{Z}_2)$ is well-defined if $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have finite first moments (see Baringhaus and Franz 2004, lem. 2.3). With the above discussion in mind, we are now in a position to define the rank-based version of the energy measure.

Definition 3.3 (Rank energy). Suppose that $\mathbf{Z}_1 \sim \mu_1$ and $\mathbf{Z}_2 \sim \mu_2$ are independent and $\mu_1, \mu_2 \in \mathcal{P}_{ac}(\mathbb{R}^d)$. Fix some $\lambda \in (0, 1)$ (prespecified). Also let $\mathbf{R}_\lambda(\cdot)$ denote the population rank map (see Definition 2.1) corresponding to the mixture distribution $\lambda\mu_1 + (1-\lambda)\mu_2$. Then the rank energy ($\mathrm{RE}_\lambda^2$) between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is defined as follows:

$$\mathrm{RE}_\lambda^2(\mathbf{Z}_1, \mathbf{Z}_2) := \gamma_d \int_{\mathbb{R}} \int_{\mathcal{S}^{d-1}} \big[\mathbb{P}(\mathbf{a}^\top\mathbf{R}_\lambda(\mathbf{Z}_1) \le t) - \mathbb{P}(\mathbf{a}^\top\mathbf{R}_\lambda(\mathbf{Z}_2) \le t)\big]^2 \, d\kappa(\mathbf{a})\, dt. \tag{13}$$

In other words, the rank energy between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is exactly equal to the usual energy measure between $\mathbf{R}_\lambda(\mathbf{Z}_1)$ and $\mathbf{R}_\lambda(\mathbf{Z}_2)$. Note that Equation (13) is well-defined without any moment assumptions.

Remark 3.5. The choice of $\lambda \in (0, 1)$ in Definition 3.3 may seem subjective. However, in the kind of applications, we are interested in, we will see that a natural choice of λ will surface from the context of the problem itself.

Now, let us inspect the properties of RE_λ^2 which make it a desirable candidate for measuring two-sample goodness of fit. The proof of the following result is given in Appendix F.5 (see the supplement).

Lemma 3.2. Under the same assumptions as in Definition 3.3, we have

(a) Suppose that $\mathbf{Z}_1^1, \mathbf{Z}_1^2$ are iid with the same distribution as $\mathbf{Z}_1$, and $\mathbf{Z}_2^1, \mathbf{Z}_2^2$ are iid with the same distribution as $\mathbf{Z}_2$. Then, $\mathrm{RE}_\lambda^2(\mathbf{Z}_1, \mathbf{Z}_2) = 2\mathbb{E}\|\mathbf{R}_\lambda(\mathbf{Z}_1^1) - \mathbf{R}_\lambda(\mathbf{Z}_2^1)\| - \mathbb{E}\|\mathbf{R}_\lambda(\mathbf{Z}_1^1) - \mathbf{R}_\lambda(\mathbf{Z}_1^2)\| - \mathbb{E}\|\mathbf{R}_\lambda(\mathbf{Z}_2^1) - \mathbf{R}_\lambda(\mathbf{Z}_2^2)\|$.
(b) $\mathrm{RE}_\lambda^2(\mathbf{Z}_1, \mathbf{Z}_2) = 0$ if and only if $\mathbf{Z}_1 \overset{d}{=} \mathbf{Z}_2$.
(c) (Invariance) Suppose that $\mathbf{a} \in \mathbb{R}^d$ and $b > 0$. Then $\mathrm{RE}_\lambda^2(\mathbf{Z}_1, \mathbf{Z}_2) = \mathrm{RE}_\lambda^2(\mathbf{a} + b\mathbf{Z}_1, \mathbf{a} + b\mathbf{Z}_2)$.
(d) Suppose that $\mathbf{Z}_1^n$ and $\mathbf{Z}_2^n$ are two independent sequences of random vectors having absolutely continuous distributions such that $\mathbf{Z}_1^n \overset{w}{\longrightarrow} \mathbf{Z}_1$ and $\mathbf{Z}_2^n \overset{w}{\longrightarrow} \mathbf{Z}_2$ as $n \to \infty$. Then, $\mathrm{RE}_\lambda^2(\mathbf{Z}_1^n, \mathbf{Z}_2^n) \longrightarrow \mathrm{RE}_\lambda^2(\mathbf{Z}_1, \mathbf{Z}_2)$ as $n \to \infty$.

4. Distribution-Free Multivariate Independence and Equality of Distributions Testing

This section is devoted to developing the new multivariate rank-based distribution-free testing procedures for the nonparametric problems discussed in the Introduction.

4.1. Distribution-Free Mutual Independence Testing

Suppose that $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ are iid observations from some probability distribution $\mu \in \mathcal{P}(\mathbb{R}^{d_1+d_2})$ (here $d_1, d_2 \geq 1$) with marginals $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$. In this subsection, we assume that

$$\textbf{(AP1)}: \qquad \mu_{\mathbf{X}} \in \mathcal{P}_{ac}(\mathbb{R}^{d_1}) \qquad \text{and} \qquad \mu_{\mathbf{Y}} \in \mathcal{P}_{ac}(\mathbb{R}^{d_2}).$$

We are interested in testing the hypothesis: $\mathrm{H}_0 : \mu = \mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}}$ versus $\mathrm{H}_1 : \mu \neq \mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}}$. The above is certainly a classical problem in statistics and has received widespread attention across many decades. One of the earliest approaches, for $d_1 = d_2 = 1$, was through the introduction of Pearson's correlation (see, e.g., Pearson 1920), which was later modified into rank-based correlation measures such as Spearman's rank correlation (see Spearman 1904) and Kendall's τ (see Kendall 1938; Kendall and Gibbons 1990). For an overview of other parametric approaches to the above problem, see Wilks (1938) and Pillai and Jayachandran (1967) and the references therein. However, nonparametric testing procedures soon replaced parametric ones as they do not require strong modeling assumptions and are consequently more robust and generally applicable.

One of the first nonparametric approaches to the above problem, when $d_1 = d_2 = 1$, was by Hoeffding (1948), where the author proposed a test based on empirical distribution functions; also see Blum, Kiefer, and Rosenblatt (1961). A "quadrant"-based procedure was introduced in the late 1950s by Mosteller (see Mosteller 1946) and later analyzed in Blomqvist (1950); also see Gieser and Randles (1997). A density estimation-based approach to independence testing was proposed in Rosenblatt (1975). When either $d_1 > 1$ or $d_2 > 1$, perhaps the most common approach historically used coordinate-wise or spatial ranks and signs (see, e.g., Puri and Sen 1971; Oja and Randles 2004; Oja 2010 and the references therein). Such coordinate-wise rank-based extensions to Spearman's rank correlation, Kendall's τ and the quadrant statistic (mentioned above) for testing independence, when $d_1 > 1$ or $d_2 > 1$, were proposed in Taskinen, Kankainen, and Oja (2003), Taskinen, Oja, and Randles (2005). In Friedman and Rafsky (1983), the authors presented a graph-based test of independence. A density-based approach, involving the estimation of mutual information has been used in Berrett and Samworth (2019). Other proposals include the use of a maximal (or total) information coefficient (see Reshef et al. 2016, 2018), empirical copula processes (see Kojadinovic and Holmes 2009; Quessy 2010), ranks of pairwise distances (see Heller, Heller, and Gorfine 2013), etc. A kernel-based method, namely the Hilbert-Schmidt Independence criteria, which perhaps dates back to 1959 (see Rényi 1959) has also been recently studied in great detail by Gretton, Fukumizu et al. (2008), Gretton, Bousquet, et al. (2005), and Gretton, Herbrich, et al. (2005). Given the huge body of work in this area, we refer the reader to Drouet Mari and Kotz (2001), Josse and Holmes (2016) for a survey on other testing procedures existing in the literature. While some of the tests discussed above guarantee consistency against fixed alternatives, a recurrent problem with all these approaches is that they lack the exact distribution-free property when either $d_1 > 1$ or $d_2 > 1$.

The only distribution-free test in the context of mutual independence testing was proposed in Heller, Gorfine, and Heller (2012); also see Biswas, Sarkar, and Ghosh (2016), Heller and Heller (2016). However, none of these tests come with any result that guarantees consistency against all fixed alternatives.

Over the past 40 years or so, multivariate tests of independence based on empirical characteristic functions have gained some prominence, thanks to early works in Kankainen (1995), Csörgő (1985), and Feuerverger (1993) and most significantly due to the seminal work by Szekely and coauthors (see Bakirov, Rizzo, and Székely 2006; Székely, Rizzo, and Bakirov 2007; Székely and Rizzo 2009), where the notion of distance covariance was introduced; recall that in Section 3.1 we have already encountered the population version of this measure. Interestingly, distance covariance can also be interpreted as a weighted integral in terms of the difference between the joint empirical characteristic function and the product of marginal characteristic functions (see Székely, Rizzo, and Bakirov 2007). Distance covariance also has interesting connections to kernel-based methods; see for example, Sejdinovic et al. (2013). On account of being simple to implement, easily explainable and providing consistency against any fixed alternatives (under suitable moment assumptions), this testing procedure has attracted a lot of attention, has inspired many applications, and is still a subject of active research.

In this section, we introduce a distribution-free multivariate rank-based version of the distance covariance test (see e.g., Székely, Rizzo, and Bakirov 2007) and demonstrate its appealing properties. We describe our method below. Let $\mu_n^{\mathbf{X}}$ and $\mu_n^{\mathbf{Y}}$ denote the empirical distributions on $\mathcal{D}_n^{\mathbf{X}} := \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ and $\mathcal{D}_n^{\mathbf{Y}} := \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$, respectively. Moreover, let $\mathcal{H}_n^{d_1} := \{\mathbf{h}_1^{d_1}, \dots, \mathbf{h}_n^{d_1}\}$ and $\mathcal{H}_n^{d_2} := \{\mathbf{h}_1^{d_2}, \dots, \mathbf{h}_n^{d_2}\}$ denote the (fixed) sample of d_1 and d_2-dimensional ranks (analogous to $\mathbf{c}_i$'s in (2)). For $i = 1, 2$, as in Section 2.2, we recommend the use of a standard d_i-dimensional quasi-Monte Carlo sequence (see Appendix E.3 in the supplement for a discussion) when $d_i > 1$ and the standard $\{i/n\}_{i \leq n}$ grid when $d_i = 1$. We will work under the following assumption on $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$:

(AP2): The empirical distributions on $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$ converge weakly to $\mathcal{U}^{d_1}$ and $\mathcal{U}^{d_2}$, respectively.

Finally, we shall use $\hat{\mathbf{R}}_n^{\mathbf{X}}(\cdot)$ and $\hat{\mathbf{R}}_n^{\mathbf{Y}}(\cdot)$ to denote the empirical rank maps (see Definition 2.2) corresponding to the transportation of $\mu_n^{\mathbf{X}}$ and $\mu_n^{\mathbf{Y}}$ to the empirical distributions on $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$, respectively (see Equation (6)). Next, we define

$$\mathrm{RdCov}_n^2 := S_1 + S_2 - 2S_3 \tag{14}$$

where $S_1 := \dfrac{1}{n^2} \displaystyle\sum_{k,l=1}^{n} \|\hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_k) - \hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_l)\| \|\hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_k) - \hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_l)\|,$

$$S_2 := \left(\frac{1}{n^2} \sum_{k,l=1}^{n} \|\hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_k) - \hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_l)\| \right)$$

$$\times \left(\frac{1}{n^2} \sum_{k,l=1}^{n} \|\hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_k) - \hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_l)\| \right),$$

$$S_3 := \frac{1}{n^3} \sum_{k,l,m=1}^{n} \|\hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_k) - \hat{\mathbf{R}}_n^{\mathbf{X}}(\mathbf{X}_l)\| \|\hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_k) - \hat{\mathbf{R}}_n^{\mathbf{Y}}(\mathbf{Y}_m)\|.$$

Observe that the right-hand side of Equation (14) can be viewed as an empirical version of population RdCov (see Equation (10)) through its alternate expression as in Lemma 3.1 (part (a)). RdCov_n^2 can also be viewed as a rank-transformed version of the empirical distance covariance measure as introduced in (Székely, Rizzo, and Bakirov 2007, eqs. (2.9) and (2.18)). By Székely, Rizzo, and Bakirov (2007, theor. 1), it is easy to see that the right-hand side of Equation (14) is always nonnegative. Moreover, note that, given the ranks, RdCov_n^2 can be computed in $\mathcal{O}(n^2(d_1 + d_2))$ steps (see Huo and Székely 2016). In the following lemma, we demonstrate the distribution-free property of RdCov_n^2 (see Appendix F.6 in the supplement for a proof).

Lemma 4.1. Under assumption **(AP1)** and H_0, the distribution of RdCov_n^2, as defined in Equation (14), is free of $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$.

Distribution-free independence testing procedure: Given a (prespecified) Type I error level $\alpha \in (0, 1)$, let $c_n := \inf\{c > 0 : \mathbb{P}_{\mathrm{H}_0}(n\mathrm{RdCov}_n^2 \geq c) \leq \alpha\}$. Note that, under H_0, RdCov_n^2 is distribution-free (by Lemma 4.1) and therefore, so is c_n. In other words, c_n depends only on $n, d_1, d_2, \mathcal{H}_n^{d_1}, \mathcal{H}_n^{d_2}$ and α, and can consequently be determined even before the data is observed. Moreover, we show in Theorem 4.1 that, if assumption *(AP2)* is satisfied then asymptotically c_n does not even depend on the particular choice of $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$. Given c_n, our proposed testing procedure rejects H_0 if $n\mathrm{RdCov}_n^2 \geq c_n$ and accepts H_0 otherwise. By definition of c_n, this is clearly a level α test.

Remark 4.1. The notion of rank-based distance covariance has attracted some interest in the literature. For $d_1 = d_2 = 1$, it has been discussed in Székely and Rizzo (2009), although to the best of our knowledge, its theoretical properties have not been analyzed. In the discussion, Rémillard (2009) based on Székely and Rizzo (2009), the author proposed using distance covariance based on the vectors of component-wise ranks (for general d_1, d_2). This idea also has connections with existing copula-based approaches; see, for example, Kojadinovic and Holmes (2009) for details. This approach however does not yield a distribution-free test (if either d_1 or d_2 is > 1), neither for finite n nor asymptotically. In that sense, our proposal provides the "correct" version of rank-based distance covariance.

Remark 4.2. The computation of RdCov_n^2 incurs a worst-case complexity of $\mathcal{O}(n^3 + n^2(d_1 + d_2))$ which is larger than $\mathcal{O}(n^2)$ complexity for kernel tests, (e.g., Gretton and Györfi (2008)) and $\mathcal{O}(kn \log n)$ for k-nearest neighbor methods (e.g., Berrett and Samworth 2019). This is a price that we are paying for getting exactly distribution-free tests. Note that our proposed procedure does not require any resampling methods to determine cutoffs unlike kernel or nearest neighbor procedures.

One of the interesting features of our proposed statistic, that is, RdCov_n^2, is that it has a close connection with the celebrated Hoeffding's D-statistic (see Hoeffding 1948)—one of the earliest nonparametric approaches to testing for mutual independence when $d_1 = d_2 = 1$. In fact, RdCov_n^2 is exactly equivalent to the statistic proposed in Blum, Kiefer, and Rosenblatt (1961) (also see the right-hand sides of Equations (15) and (16) for the population and the empirical versions, respectively), which in turn is a modified version of Hoeffding's D-statistic. The following lemma (see Appendix F.7 in the supplement for a proof) makes this connection precise (also see Weihs, Drton, and Meinshausen 2018).

Lemma 4.2. Suppose that $(X, Y) \in \mathbb{R}^2$ with bivariate distribution function (DF) $F^{X,Y}(\cdot)$, and corresponding marginal DFs, F^X and F^Y. Assume that $F^X(\cdot)$ and $F^Y(\cdot)$ are absolutely continuous. Also suppose that random samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ are drawn according to the same distribution as (X, Y). Furthermore, we will use $F_n^{X,Y}(\cdot)$, $F_n^X(\cdot)$ and $F_n^Y(\cdot)$ to denote the joint and marginal empirical DFs of X_i's and Y_i's, respectively. Then, the following holds:

$$\frac{1}{4}\mathrm{RdCov}^2(X, Y) = \int_{\mathbb{R}^2} \left(F^{X,Y}(x, y) - F^X(x)F^Y(y)\right)^2 dF^X(x)\, dF^Y(y) \quad \text{and,} \tag{15}$$

$$\frac{1}{4}\mathrm{RdCov}_n^2 = \int \left(F_n^{X,Y}(x, y) - F_n^X(x)F_n^Y(y)\right)^2 dF_n^X(x)\, dF_n^Y(y). \tag{16}$$

We are now interested in two fundamental questions about our proposed test: (a) "What is the limiting distribution of our test statistic?"; (b) "Is our test consistent against all fixed alternatives, as the sample size grows?". We investigate these two questions in Theorems 4.1 and 4.2, respectively (see Appendices F.8 and F.9 in the supplement for the proofs).

Theorem 4.1. Under assumptions **(AP1)**, **(AP2)** and under H_0, there exists universal nonnegative constants $(\eta_1, \eta_2, \ldots)$ such that $n\mathrm{RdCov}_n^2 \xrightarrow{w} \sum_{j=1}^{\infty} \eta_j Z_j^2$ as $n \to \infty$, where $Z_1, Z_2, \ldots$ are iid standard Gaussian random variables. In fact, η_j's do not depend on the specific choice of $\mathcal{H}_n^{d_1}$ or $\mathcal{H}_n^{d_2}$ as long as **(AP2)** is satisfied.

Remark 4.3 (Limiting distribution). The limiting distribution in Theorem 4.1 is exactly the same as that of usual distance covariance (under H_0) when $\mu_{\mathbf{X}} = \mathcal{U}^{d_1}$ and $\mu_{\mathbf{Y}} = \mathcal{U}^{d_2}$ (see (Székely, Rizzo, and Bakirov 2007, theor. 5)).

Remark 4.4 (Distribution-freeness). Note that the asymptotic distribution of the usual distance covariance statistic, given in (Székely, Rizzo, and Bakirov 2007, theor. 5), depends on $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$, which are unknown. As a result, even for large n, in practice, one usually has to resort to resampling/permutation techniques or further worst case approximations (see Székely, Rizzo, and Bakirov 2007, theor. 6) to determine the critical value of the test. Having finite sample (and asymptotic) distribution-freeness avoids the need for such approximation techniques (for small as well as large n). In Appendix D.4 (see the supplement), we discuss (computationally) how large n should be (depending on d_1 and d_2) so as to use quantiles from the asymptotic distribution of $n\text{RdCov}_n^2$ to approximate thresholds for our testing procedure (see Tables D.4–D.7 in the supplement). In Table D.8 (in Appendix D.4 in the supplement), we provide the universal asymptotic 0.95-quantiles as d_1, d_2 varies (for $d_1, d_2 \le 8$).

Remark 4.5 (Our proof technique). Observe that, contrary to the study of the usual distance covariance Székely, Rizzo, and Bakirov (2007) which can be analyzed using standard techniques from empirical process theory (as in Székely, Rizzo, and Bakirov 2007, theor. 5) or results from degenerate V-statistics (as used in (Lyons 2013, theor. 2.7)), the study of RdCov_n^2 is more complicated as it involves dependent multivariate ranks. Our main technique for proving Theorem 4.1 is to use Hoeffding's combinatorial central limit theorem (see, e.g., Chen and Fang 2015). In the process, we prove some results on permutation statistics (see Lemma F.1 in the supplement) which may be of independent interest.

The following result (proved in Appendix F.9 in the supplement) shows that our proposed testing procedure yields a consistent sequence of tests under fixed alternatives (i.e., the power of our test converges to 1, as the sample size increases, for any fixed alternative).

Theorem 4.2. Under assumptions (AP1) and (AP2), $\text{RdCov}_n^2 \xrightarrow{\text{a.s.}} \text{RdCov}^2(\mathbf{X}, \mathbf{Y})$ as $n \to \infty$, where $(\mathbf{X}, \mathbf{Y}) \sim \mu$. Moreover, $\mathbb{P}(n\text{RdCov}_n^2 > c_n) \longrightarrow 1$, as $n \to \infty$, provided $\mu \neq \mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}}$.

Remark 4.6 (Minimal assumptions). The proof of Theorem 4.2 reveals that only the a.s.-convergence of empirical transport maps in the L^2-norm (see Theorem 2.1) is necessary. Therefore, by resorting to a weaker form of convergence (as compared to the L^∞-convergence as in Chernozhukov et al. 2017; Ghosal and Sen 2019; Hallin et al. 2021) we have effectively reduced the set of assumptions needed on $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$ for getting a consistent sequence of tests (contrary to Ghosal and Sen 2019). Moreover we are able to establish consistency without any moment assumptions (contrary to (Székely, Rizzo, and Bakirov 2007, theor. 2)).

Remark 4.7 (Quasi-Monte Carlo sequence). Corollary E.1 ensures that assumption *(AP2)* is satisfied for quasi-Monte Carlo sequences (see Appendix E.3 in the supplement for details and examples).

Remark 4.8 (Invariance under coordinate-wise monotone transformations). An alternate approach to testing mutual independence would be to transform the observed data into their marginal one-dimensional ranks first and then construct the multivariate ranks based on this transformed data. Let us elaborate on this briefly. Let us write $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id_1})$ in terms of its univariate components, for $1 \le i \le n$. For $1 \le j \le d_1$, construct $\tilde{\mathbf{X}}_i$ such that $\tilde{X}_{ij}$ equals the usual one-dimensional rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$. Repeat the same exercise with the $\mathbf{Y}_i$'s to form $\tilde{\mathbf{Y}}_i$'s. Now, consider $\{(\tilde{\mathbf{X}}_i, \tilde{\mathbf{Y}}_i)\}_{i=1}^n$ and obtain multivariate ranks of $\tilde{\mathbf{X}}_i$'s and $\tilde{\mathbf{Y}}_i$'s using measure transportation as described above (see Equation (7)). Finally, calculate a suitable test statistic for independence (such as RdCov_n^2) based on these ranks. This approach has natural connections to copula-based methods (see Kojadinovic and Holmes 2009) and ensures that the constructed tests will be invariant under coordinate-wise monotone transformations of the data (see Lemma 3.1, part (d)). An analogous theoretical analysis can be carried out for this modified procedure.

4.2. Distribution-Free Multivariate Two-Sample Testing

Here, we shall consider the two-sample goodness-of-fit testing problem in a multivariate setting. Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_m \overset{\text{iid}}{\sim} \mu_{\mathbf{X}}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{\text{iid}}{\sim} \mu_{\mathbf{Y}}$ (independent of the $\mathbf{X}_i$'s), where we assume that

$$\textbf{(AP3)}: \qquad \mu_{\mathbf{X}}, \mu_{\mathbf{Y}} \in \mathcal{P}_{ac}(\mathbb{R}^d).$$

We are interested in testing the hypothesis: $\mathrm{H}_0 : \mu_{\mathbf{X}} = \mu_{\mathbf{Y}}$ versus $\mathrm{H}_1 : \mu_{\mathbf{X}} \neq \mu_{\mathbf{Y}}$. The two-sample problem (or its multi-sample extension) has been studied in great detail over the years. In this context, rank and data-depth-based methods have mostly been restricted to testing against location-scale alternatives, see, for example, Hettmansperger, Möttönen, and Oja (1998), Randles and Peters (1990), and Möttönen and Oja (1995). Distribution-free depth-based tests which are consistent if restricted to the above class of alternatives are discussed in Liu et al. (2010) and Rousson (2002). An alternative route for testing against general alternatives includes graph-based tests such as in Friedman and Rafsky (1979), where the authors constructed a test based on the minimum spanning tree of a graph with the data points as its vertices and pairwise distances as edge weights. Various interesting modifications and extensions to this test have been proposed in literature, see, for example, Chen and Friedman (2017), Henze (1988), Schilling (1986), and Petrie (2016). Theoretical properties of all these tests can be studied under a unified framework as shown in Bhattacharya (2019).

As mentioned in the Introduction, there are only a few multivariate nonparametric distribution-free two-sample goodness-of-fit tests. In Rosenbaum (2005) (also see Arias-Castro and Pelletier 2016; Agarwal et al. 2019 for subsequent theoretical analysis), Rosenbaum constructed his distribution-free test statistic from a minimum non-bipartite matching (see Lu et al. 2011) of the pooled sample of observations.

The recent article (Boeckel, Spokoiny, and Suvorikova 2018) proposed a distribution-free two-sample goodness-of-fit test using multivariate ranks (based on optimal transport). Thus, their approach is quite similar to ours. However, the authors constructed the empirical ranks by replacing the $\mathbf{h}_i$'s (fixed quasi-Monte Carlo sequence) in Equation (7) with a random draw of n iid uniforms which makes the test statistic random, given the data (due to the external randomization). Further, in Boeckel, Spokoiny, and Suvorikova (2018), the authors proposed a test for the two-sample equality of distributions using the Wasserstein distance, instead of the energy statistic that we use. Their theoretical results require much stronger assumptions on the underlying distributions, for example, the assumption that the data generating distribution be compactly supported in addition to being absolutely continuous. Moreover, the authors do not develop the asymptotic null distribution theory of their proposed test.

Yet another class of pairwise-distance-based tests use ideas from reproducing kernel Hilbert spaces (RKHS), see, for example, Gretton et al. (2009, 2012). The principle idea here is to embed probability distributions in RKHSs through what are called *mean embeddings* and measure goodness of fit between two distributions by the Hilbert–Schmidt norm between the corresponding mean embeddings. These kernel-based measures can alternatively be expressed as *probability integral metrics* which equal 0 if and only if the underlying distributions are exactly the same. In fact, the energy statistic (see Baringhaus and Franz 2004; Székely and Rizzo 2013)—a popular and powerful goodness-of-fit measure—can also be viewed as a special case of kernel-based methods (see Sejdinovic et al. 2013). Due to its simplicity, the energy distance has been studied and applied extensively over the past decade, as we have already highlighted in Section 1. However, note that a common disadvantage of these kernel-based methods (including the usual energy statistic) is that they are not exactly distribution-free.

In this subsection, we propose the rank energy statistic—a distribution-free goodness-of-fit measure based on the energy distance—for testing the equality of two multivariate distributions. We describe our method below. We will use $\mu_m^{\mathbf{X}}$ and $\mu_n^{\mathbf{Y}}$ to denote the empirical distributions on $\mathcal{D}_m^{\mathbf{X}} := \{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$ and $\mathcal{D}_n^{\mathbf{Y}} := \{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$, respectively. Let $\mu_{m,n}^{\mathbf{X},\mathbf{Y}} := (m+n)^{-1}(m\mu_n^{\mathbf{X}} + n\mu_n^{\mathbf{Y}})$ and let $\mathcal{H}_{m+n}^d := \{\mathbf{h}_1^d, \ldots, \mathbf{h}_{m+n}^d\} \subset [0,1]^d$ denote the (fixed) sample multivariate ranks. We will further work under the following assumption on $\mathcal{H}_{m+n}^d$:

(AP4) The empirical distribution on $\mathcal{H}_{m+n}^d$ converges weakly to $\mathcal{U}^d$ as $\min(m,n) \to \infty$. Note that choosing $\mathcal{H}_{m+n}^d$ to be a d-dimensional quasi-Monte Carlo sequence, for $d \geq 2$, and $\{i/(m+n) : 1 \leq i \leq m+n\}$ for $d = 1$, ensures that **(AP4)** is satisfied (see Corollary E.1 for details).

Finally, we shall use $\hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\cdot)$ to denote the joint empirical rank map (see Definition 2.2) corresponding to the transportation of $\mu_{m,n}^{\mathbf{X},\mathbf{Y}}$ to the empirical distribution on $\mathcal{H}_{m+n}^d$. The rank energy statistic is defined as follows:

$$\begin{aligned}\mathrm{RE}_{m,n}^2 := \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|\hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{X}_i) - \hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{Y}_j)\| \\ - \frac{1}{m^2}\sum_{i,j=1}^{m}\|\hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{X}_i) - \hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{X}_j)\| \\ - \frac{1}{n^2}\sum_{i,j=1}^{n}\|\hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{Y}_i) - \hat{\mathbf{R}}_{m,n}^{\mathbf{X},\mathbf{Y}}(\mathbf{Y}_j)\|. \qquad (17)\end{aligned}$$

Observe that the right-hand side of Equation (17) can be viewed as an empirical version of RE^2 (see (13)) through its alternate expression as in Lemma 3.2 (part (a)). $\mathrm{RE}_{m,n}^2$ can also be viewed as a rank-transformed version of the empirical energy measure as in (Székely and Rizzo 2013, eq. (6.1)). Due to space constraints, we will refer the interested reader to Székely and Rizzo (2013) for further motivation of the energy statistic. By Baringhaus and Franz (2004, eq. (5)), it is easy to see that the right-hand side of Equation (17) is always nonnegative. Just as RdCov_n^2 (in Equation (14)), $\mathrm{RE}_{m,n}^2$ above can also be computed in $\mathcal{O}(mnd)$ steps (see Zhao and Meng 2015) given the vector of multivariate ranks. In the following lemma, we illustrate the distribution-free property of $\mathrm{RE}_{m,n}^2$ (see Appendix F.10 in the supplement for a proof).

Lemma 4.3. Under assumption (*AP3*) and under H_0, the distribution of $\mathrm{RE}_{m,n}^2$, as defined in Equation (17), is free of $\mu_{\mathbf{X}} \equiv \mu_{\mathbf{Y}}$.

Distribution-free two-sample testing procedure: Given a (prespecified) Type I error level $\alpha \in (0,1)$, let $c_{m,n} := \inf\{c > 0 : \mathbb{P}_{\mathrm{H}_0}(mn(m+n)^{-1}\mathrm{RE}_{m,n}^2 \geq c) \leq \alpha\}$. As $\mathrm{RE}_{m,n}^2$ is distribution-free under H_0 (by Lemma 4.3), so is $c_{m,n}$. Given $c_{m,n}$, our proposed testing procedure rejects H_0 if $mn(m+n)^{-1}\mathrm{RE}_{m,n}^2 \geq c_{m,n}$ and accepts H_0 otherwise. This results in a level α test.

An interesting feature of our proposed statistic $\mathrm{RE}_{m,n}^2$ is its equivalence with the celebrated Cramér-von Mises statistic for two sample equality of distributions testing (see e.g., Anderson (1962) and the right-hand side of Equation (18)) when $d = 1$. The following lemma (see Appendix F.11 in the supplement for a proof) makes this connection precise.

Lemma 4.4. For $d = 1$, let F_m^X, G_n^Y and $H_{m+n}^{X,Y}$ denote the empirical distribution functions on $\{X_1, \ldots, X_m\}$, $\{Y_1, \ldots, Y_n\}$ and the pooled sample, respectively. Then,

$$\frac{1}{2}\,\mathrm{RE}_{m,n}^2 = \int \left(F_m^X(t) - G_n^Y(t)\right)^2 dH_{m+n}^{X,Y}(t). \qquad (18)$$

The right-hand side of Equation (18) is the exact Cramér-von Mises statistic as in Anderson (1962). At the population level, fix any $\lambda \in (0,1)$ and let F^X, G^Y and $H_\lambda^{X,Y}$ be the distribution functions associated with the probability measures μ_X, μ_Y and $\lambda\mu_X + (1-\lambda)\mu_Y$. Assume also that F^X and G^Y are absolutely continuous. Then,

$$\frac{1}{2}\mathrm{RE}_\lambda^2(X,Y) = \int_{-\infty}^{\infty}\left(F^X(t) - G^Y(t)\right)^2 dH_\lambda^{X,Y}(t). \qquad (19)$$

Next we find the asymptotic distribution of $\mathrm{RE}_{m,n}^2$ in Theorem 4.3 and prove the consistency of our proposed procedure in Theorem 4.4; see Appendices F.8 and F.13 in the supplement for their proofs.

Theorem 4.3. Suppose that $\min(m,n) \to \infty$. Under assumptions (AP3), (AP4) and under H_0, we have $\frac{mn}{m+n}\mathrm{RE}_{m,n}^2 \xrightarrow{w} \sum_{j=1}^{\infty}\tau_j Z_j^2$ as $n \to \infty$, where $Z_1, Z_2, \ldots$ are iid standard normals and τ_j's are fixed nonnegative constants. In fact, τ_j's do not depend on the specific choice of $\mathcal{H}_{m+n}^d$ as long as (AP4) is satisfied.

Remark 4.9 (Limiting distribution). The limiting distribution in Theorem 4.3 is exactly the same as that of the usual energy statistic (under H_0) when $\mu_{\mathbf{X}} = \mu_{\mathbf{Y}} = \mathcal{U}^d$ (Baringhaus and Franz 2004, theor. 2.3).

Theorem 4.4. Suppose that $m/(m+n) \longrightarrow \lambda \in (0,1)$. Then, under assumptions (*AP3*) and (*AP4*), $\mathrm{RE}_{m,n}^2 \xrightarrow{a.s.} \mathrm{RE}_\lambda^2(\mathbf{X},\mathbf{Y})$ as $n \to \infty$, where $\mathbf{X} \sim \mu_{\mathbf{X}}$ and $\mathbf{Y} \sim \mu_{\mathbf{Y}}$ (note the connection with Remark 3.5). Moreover, $\mathbb{P}(mn(m+n)^{-1}\mathrm{RE}_{m,n}^2 > c_n) \longrightarrow 1$, as $n \to \infty$, provided $\mu_{\mathbf{X}} \neq \mu_{\mathbf{Y}}$.

The multivariate two-sample testing procedure described above bears all the useful properties of our independence testing procedure from Section 4.1. In particular, the proposed test is distribution-free for each fixed m and n and also in an asymptotic sense. In Appendix D.4 (see the supplement), we study, using simulations, how large m, n should be (depending on d) so as to reasonably use quantiles from the asymptotic distribution of $mn(m+n)^{-1}\mathrm{RE}^2_{m,n}$ to determine thresholds for our testing procedure (see Tables D.9–D.12 in the supplement). In Table D.13 (in Appendix D.4 in the supplement), we provide universal asymptotic quantiles (5%) up to $d \le 8$.

Our proposed test is also consistent against fixed alternatives without any moment assumptions, as opposed to the usual test based on the energy statistic (see Baringhaus and Franz 2004; Székely and Rizzo 2013). Moreover, we are also able to reduce the smoothness assumptions on the underlying measures $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$ necessary for consistency (cf. Ghosal and Sen 2019, prop. 5.2; Boeckel, Spokoiny, and Suvorikova 2018, theor. 3.1).

4.3. *Extensions to the K-Sample Problem*

The methods we discussed in Sections 4.1 and 4.2 have natural extensions to the K-sample setting; namely, testing for mutual independence of K random vectors, and multivariate goodness-of-fit testing for K populations (as mentioned in Section 1). Using the same principles as above, we can again construct exact distribution-free tests for the above problems that will be consistent against all fixed alternatives. Due to space constraints, we relegate a detailed discussion of this to Appendix E; in particular, see Propositions E.1 and E.2 (in the supplement).

5. Discussion and Recent Developments

We have developed a framework for multivariate distribution-free nonparametric testing using the method of multivariate ranks defined using the theory of optimal transportation (motivated from Hallin 2017; Hallin et al. 2021). We have illustrated our general approach through two problems: (I) testing for mutual independence of K ($\geq$ 2) random vectors, and (II) goodness-of-fit testing for K ($\geq$ 2) multivariate distributions. We show that our proposed tests are finite sample distribution-free, consistent against all alternatives (under minimal assumptions), and are computationally feasible. In fact, the proposed tests reduce to well-known one-dimensional tests for problems (I) and (II). We further derive the asymptotic weak limits of our test statistics, under the null hypotheses. In the process, we also derive results on the asymptotic regularity of optimal transport maps (aka multivariate ranks) which is of independent interest. As far as we are aware, this is the first attempt to systematically develop distribution-free multivariate tests that are consistent against all alternatives and are computationally feasible.

Motivated by our framework, a number of articles have subsequently been written in this active research area in the last one and a half years since the first `arxiv` version of our article was posted (as mentioned in Section 1, the article Shi, Drton, and Han (2020) was posted a couple of days after ours). Some of these include Hallin, La Vecchia, and Liu 2020a,b (in time series applications), Shi et al. 2020; Deb, Ghosal, and Sen 2020 (in independence testing), and Deb, Bhattacharya, and Sen 2021 (in goodness-of-fit testing). In fact, these articles generalize our strategy further by fixing a reference distribution (as we do with Unif$[0,1]^d$ as the reference distribution) and proposing statistics based on score/kernel function-based transforms of the obtained multivariate ranks (see, e.g., Shi et al. 2020, eq. 4.2), leading to a larger class of tests. The most general approach in this regard came recently in Deb, Bhattacharya, and Sen (2021) where the authors allowed for flexible reference distributions in addition to general score functions.

From a theoretical perspective, the past year has seen a number of important advancements in nonparametric distribution-free testing using optimal transport. The first Hájek representation (see Hájek and Šidák 1967) result for a class of score-function transformed multivariate rank statistics, under appropriate null hypothesis, was developed in Shi et al. (2020, theor. 5.1). Similar asymptotic representations for other statistics have since been developed in Deb, Ghosal, and Sen (2020), Deb, Bhattacharya, and Sen (2021), Hallin, La Vecchia, and Liu (2020a), and Hallin, La Vecchia, and Liu (2020b). The standing problem of consistency of such tests under general fixed alternatives has recently been resolved in Deb, Bhattacharya, and Sen (2021, prop. 4.2). Further, in Deb, Bhattacharya, and Sen (2021), the authors showed that the asymptotic relative efficiency under local alternatives of certain multivariate-rank-based tests against the Hotelling T^2 test (see Hotelling 1931) for goodness of fit is *at least* 1, across a large family of multivariate probability distribution, thereby making a strong case in favor of such rank-based tests. Their work provides the first multivariate, exactly distribution-free analogs of the classical results in Hodges and Lehmann (1956) and Chernoff and Savage (1958).

Supplementary Materials

The supplementary material, which is available online, contains proofs of our main results, real data experiments and additional computational studies.

Acknowledgments

We would like to thank the associate editor and the two anonymous reviewers for their constructive comments that helped improve the quality of this article.

Funding

This work was supported by NSF (grant no. DMS-2015376).

References

Agarwal, D., Mukherjee, S., Bhattacharya, B. B., and Zhang, N. R. (2019), "Distribution-Free Multisample Test Based on Optimal Matching With Applications to Single Cell Genomics," arXiv: 1906.04776. [202]

Alexandroff, A. D. (1939), "Almost Everywhere Existence of the Second Differential of a Convex Function and Some Properties of Convex Surfaces Connected With It," *Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser.* 6, 3–35. [196]

Anderson, T. W. (1962), "On the Distribution of the Two-Sample Cramer-von Mises Criterion," *Annals of Mathematical Statistics*, 33, 1148–1159. [194,203]

Arias-Castro, E., and Pelletier, B. (2016), "On the Consistency of the Crossmatch Test," *Journal of Statistical Planning and Inference*, 171 ,184–190. [202]

Bakirov, N. K., Rizzo, M. L., and Székely, G. J. (2006), "A Multivariate Nonparametric Test of Independence," *Journal of Multivariate Analysis*, 97, 1742–1756. [194,200]

Baringhaus, L., and Franz, C. (2004), "On a New Multivariate Two-Sample Test," *Journal of Multivariate Analysis*, 88, 190–206. [194,199,203,204]

Beneš, V., Lechnerová, R., Klebanov, L., Slámová, M., and Sláma, P. (2009), "Statistical Comparison of the Geometry of Second-Phase Particles," *Materials Characterization*, 60, 1076–1081. [194]

Berrett, T. B., and Samworth, R. J. (2019), "Nonparametric Independence Testing Via Mutual Information," *Biometrika*, 106, 547–566. [194,200,201]

Bertsekas, D. P. (1988), "The Auction Algorithm: A Distributed Relaxation Method for The Assignment Problem," *Annals of Operations Research*, 14, 105–123. [193,197]

Bhattacharya, B. B. (2019), "A General Asymptotic Framework for Distribution-Free Graph-Based Two-Sample Tests," *Journal of the Royal Statistical Society*, Series B, 81, 575–602. [202]

Biau, G., Györfi, L. (2007), "On a l_1-test Statistic of Homogeneity," *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 13, 877–881. [195]

Bickel, P. J. (1965), "On Some Asymptotically Nonparametric Competitors of Hotelling's T^2," *Annals of Mathematical Statistics*, 36, 160–173; correction, ibid. 1583. [193]

——— (1968), "A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case," *Annals of Mathematical Statistics*, 40, 1–23. [192]

Biswas, M., Mukhopadhyay, M., and Ghosh, A. K. (2014), "A Distribution-Free Two-Sample Run Test Applicable to High-Dimensional Data," *Biometrika*, 101, 913–926. [195]

Biswas, M., Sarkar, S., and Ghosh, A. K. (2016), "On Some Exact Distribution-Free Tests of Independence Between Two Random Vectors of Arbitrary Dimensions," *Journal of Statistical Planning and Inference*, 175, 78–86. [194,195,200]

Blomqvist, N. (1950), "On a Measure of Dependence Between Two Random Variables," *Annals of Mathematical Statistics*, 21, 593–600. [200]

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961), "Distribution Free Tests of Independence Based on the Sample Distribution Function," *Annals of Mathematical Statistics*, 32, 485–498. [194,200,201]

Boeckel, M., Spokoiny, V., and Suvorikova, A. (2018), "Multivariate Brenier Cumulative Distribution Functions and Their Application to Non-Parametric Testing," arXiv: 1809.04090. [193,195,196,202,204]

Brenier, Y. (1991), "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions," *Communications on Pure and Applied Mathematics*, 44, 375–417. [196]

Caffarelli, L. A. (1990), Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation. *Annals of Mathematics, (2)*, 131, 135–150. [196]

Chaudhuri, P. (1996), "On a Geometric Notion of Quantiles for Multivariate Data," *Journal of the American Statistical Association*, 91, 862–872. [193]

Chen, H., and Friedman, J. H. (2017), "A New Graph-Based Two-Sample Test for Multivariate and Object Data," *Journal of the American Statistical Association*, 112, 397–409. [202]

Chen, L. H. Y., and Fang, X. (2015), "On the Error Bound in a Combinatorial Central Limit Theorem," *Bernoulli*, 21, 335–359. [202]

Chernoff, H., and Savage, I. R. (1958), "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," *Annals of Mathematical Statistics*, 29, 972–994. [204]

Chernozhukov, V., Galichon, A., Hallin, M., Henry, M. et al. (2017), "Monge–Kantorovich Depth, Quantiles, Ranks and Signs," *The Annals of Statistics*, 45, 223–256. [193,197,202]

Conradsen, K., Nielsen, A. A., Schou, J., and Skriver, H. (2003), "A Test Statistic in the Complex Wishart Distribution and Its Application to Change Detection in Polarimetric SAR Data," *IEEE Transactions on Geoscience and Remote Sensing* 41, 4–19. [192]

Csörgő, S. (1985), "Testing for Independence by the Empirical Characteristic Function," *Journal of Multivariate Analysis*, 16, 290–299. [200]

De Philippis, G., and Figalli, A. (2013), "$W^{2,1}$ Regularity for Solutions of the Monge-Ampère Equation," *Inventiones Mathematicae*, 192, 55–69. [196]

Deb, N., Bhattacharya, B. B., and Sen, B. (2021), "Efficiency Lower Bounds for Distribution-Free Hotelling-Type Two-Sample Tests Based on Optimal Transport," arXiv: 2104.01986. [204]

Deb, N., Ghosal, P., and Sen, B. (2020), "Measuring Association on Topological Spaces Using Kernels and Geometric Graphs," arXiv: 2010.01768. [204]

Dishion, T. J., Capaldi, D. M., and Yoerger, K. (1999), "Middle Childhood Antecedents to Progressions in Male Adolescent Substance Use: An Ecological Analysis of Risk and Protection," *Journal of Adolescent Research*, 14, 175–205. [192]

Drouet Mari, D., and Kotz, S. (2001), *Correlation and Dependence*, London; Imperial College Press. Distributed by World Scientific Publishing Co., Inc. River Edge, NJ. [200]

Farris, K. B., and Schopflocher, D. P. (1999), "Between Intention and Behavior: An Application of Community Pharmacists' Assessment of Pharmaceutical Care," *Social Science & Medicine*, 49, 55–66. [192]

Feuerverger, A. (1993), "A Consistent Test for Bivariate Dependence," *International Statistical Review/Revue Internationale de Statistique*, 419–433. [200]

Folkes, V. S., Koletsky, S., and Graham, J. L. (1987), "A Field Study of Causal Inferences and Consumer Reaction: The View From the Airport," *Journal of Consumer Research*, 13, 534–539. [192]

Friedman, J. H., and Rafsky, L. C. (1979), "Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests," *Annals of Statistics*, 7, 697–717. [202]

Friedman, J. H., and Rafsky, L. C. (1983), "Graph-Theoretic Measures of Multivariate Association and Prediction," *Annals of Statistics*, 11, 377–391. [194,200]

Ghosal, P., and Sen, B. (2019), "Multivariate Ranks and Quantiles Using Optimal Transportation and Applications to Goodness-of-Fit Testing," arXiv:1905.05340. [197,202,204]

Gibbons, J. D., and Chakraborti, S. (2011), *Nonparametric Statistical Inference*, 5th ed. Statistics: Textbooks and Monographs, Boca Raton, FL: CRC Press. [193]

Gieser, P. W., and Randles, R. H. (1997), "A Nonparametric Test of Independence Between Two Vectors," *Journal of American Statistical Association*, 92, 561–567. [192,194,200]

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012), "A Kernel Two-Sample Test," *Journal of Machine Learning Research*, 13, 723–773. [195,203]

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005), "Measuring Statistical Dependence With Hilbert–Schmidt Norms," in *Algorithmic Learning Theory*, Vol. 3734 of Lecture Notes in Comput. Sci., Berlin: Springer, pp. 63–77. [200]

Gretton, A., Fukumizu, K., Harchaoui, Z., and Sriperumbudur, B. K. (2009), "A Fast, Consistent Kernel Two-Sample Test," in *Advances in Neural Information Processing Systems*, eds. Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Red Hook, NY: Curran Associates, Inc., pp. 673–681. [203]

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008), "A Kernel Statistical Test of Independence," in *Advances in Neural Information Processing Systems*, eds. J. Platt, D. Koller, Y. Singer, and S. Roweis, Red Hook, NY: Curran Associates, Inc., pp. 585–592. [194,195,200]

Gretton, A., and L. Györfi (2008), "Nonparametric Independence Tests: Space Partitioning and Kernel Approaches," in *Algorithmic Learning Theory*, eds. Yoav Freund, László G. György Turán, T. Zeugmann, Vol. 5254 of Lecture Notes in Comput. Sci. Berlin: Springer, pp. 183–198. [195,201]

Gretton, A., and L. Györfi (2010), "Consistent Nonparametric Tests of Independence," *Journal of Machine Learning Research* 11, 1391–1423. [195]

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005), "Kernel Methods for Measuring Independence," *Journal of Machine Learning Research*, 6, 2075–2129. [200]

Hájek, J. and Z. Šidák (1967). *Theory of Rank Tests*, New York: Academic Press/Prague: Academia Publishing House of the Czechoslovak Academy of Sciences. [204]

Hallin, M. (2017), "On Distribution and Quantile Functions, Ranks and Signs in $\mathbb{R}^d$," ECARES Working Papers. [193,197,204]

Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2021), "Distribution and Quantile Functions, Ranks and Signs in Dimension d: A Measure Transportation Approach," *The Annals of Statistics*, 49, 1139 – 1165. [193,195,196,197,202,204]

Hallin, M., La Vecchia, D., and Liu, H. (2020a), "Center-Outward R-Estimation for Semiparametric VARMA Models," *Journal of the American Statistical Association*, 1–14, DOI: 10.1080/01621459.2020.1832501.. [204]

——— (2020b), Rank-based testing for semiparametric VAR models: a measure transportation approach. arXiv preprint arXiv:2011.06062. [204]

Hallin, M., and Paindaveine, D. (2004), "Rank-Based Optimal Tests of the Adequacy of an Elliptic VARMA Model," *Annals Statistics*, 32, 2642–2678. [193]

——— (2006), "Parametric and Semiparametric Inference for Shape: The Role of the Scale Functional," *Statistical Decisions*, 24, 327–350. [193]

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952), *Inequalities*, 2nd ed. Cambridge: University Press. [193]

Hauke, J., and Kossowski, T. (2011), "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data," *Quaestiones Geographicae*, 30, 87–93. [198]

Heller, R., Gorfine, M., and Heller, Y. (2012), "A Class of Multivariate Distribution-Free Tests of Independence Based on Graphs," *Journal of Statistical Planning and Inference*, 142, 3097–3106. [195,200]

Heller, R., and Heller, Y. (2016), "Multivariate Tests of Association Based on Univariate Tests," in *Advances in Neural Information Processing Systems* (Vol. 29), eds. D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Red Hook, NY: Curran Associates, Inc; pp. 208–216. [195,200]

Heller, R., Heller, Y., and Gorfine, M. (2013), "A Consistent Multivariate Test of Association Based on Ranks of Distances," *Biometrika*, 100, 503–510. [194,195,200]

Henze, N. (1988), "A Multivariate Two-Sample Test Based on the Number of Nearest Neighbor Type Coincidences," *Annals of Statistics*, 16, 772–783. [202]

Hettmansperger, T. P., Möttönen, J., and Oja, H. (1998), "Affine Invariant Multivariate Rank Tests for Several Samples," *Statistical Sinica*, 8, 785–800. [202]

Hodges, Jr., J. L., and Lehmann, E. L. (1956), "The Efficiency of Some Nonparametric Competitors of the *t*-test," *Annals of Mathematical Statistics*, 27, 324–335. [204]

Hoeffding, W. (1948). "A Non-Parametric Test of Independence," *Annals of Mathematical Statistics*, 19, 546–557. [193,194,200,201]

Hollander, M., Wolfe, D. A., and Chicken, E. (2014), *Nonparametric Statistical Methods* (3rd ed.), Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley. [192]

Hotelling, H. (1931), "The Generalization of Student's Ratio," *Annals of Mathematical Statistics.*, 2, 360–378. [204]

Huo, X., and Székely, G. J. (2016), "Fast Computing for Distance Covariance," *Technometrics*, 58, 435–447. [201]

Iman, R. L., and Conover, W.-J. (1982), "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics – Simulation and Computation*, 11, 311–334. [198]

Josse, J., and Holmes, S. (2016), "Measuring Multivariate Association and Beyond," *Statistical Survey*, 10, 132–167. [200]

Kankainen, A.-L. (1995), "Consistent Testing of Total Independence Based on the Empirical Characteristic Function". Thesis (D.Phil.)–Jyvaskylan Yliopisto (Finland). Ann Arbor, MI: ProQuest LLC. . [200]

Kendall, M., and Gibbons, J. D. (1990), *Rank Correlation Methods* (5th ed.), London: Edward Arnold. [193,200]

Kendall, M. G. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30, 81–93. [200]

Klebanov, L. B. (2002), "A Class of Probability Metrics and Its Statistical Applications," in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, ed. Dodge, Y. Basel: Springer; pp. 241–252. [194]

Kojadinovic, I., and Holmes, M. (2009), "Tests of Independence Among Continuous Random Vectors Based on Cramér-von Mises Functionals of the Empirical Copula Process," *Journal Multivariate Analysis*, 100, 1137–1154. [200,201,202]

Kruskal, W. H. (1952), "A Nonparametric Test for the Several Sample Problem," *Annals of Mathematics Statistics*, 23, 525–540. [193]

Liu, J. Z., Mcrae, A. F., Nyholt, D. R., Medland, S. E., Wray, N. R., Brown, K. M., Hayward, N. K., Montgomery, G. W., Visscher, P. M., Martin, N. G., Macgregor, S. (2010), "A Versatile Gene-Based Test for Genome-Wide Association Studies," *The American Journal of Human Genetics*, 87, 139–145. [192,202]

Liu, R. Y., and Singh, K. (1993), "A Quality Index Based on Data Depth and Multivariate Rank Tests," *Journal of American Statistical Association*, 88, 252–260. [193]

Lu, B., Greevy, R., Xu, X., and Beck, C. (2011), "Optimal Nonbipartite Matching and Its Statistical Applications," *The American Statistician*, 65, 21–30. [202]

Lu, C.-J., T.-S. Lee, and C.-C. Chiu (2009) "Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression," *Decision Support Systems*, 47, 115–125. [192]

Lyons, R. (2013), "Distance Covariance in Metric Spaces," *Annals of Probability*, 41, 3284–3305. [202]

Mann, H. B., and Whitney, D. R. (1947), "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other," *Annals of Mathematical Statistics*, 18, 50–60. [193]

Marden, J. I. (1999), "Multivariate Rank Tests," in *Multivariate Analysis, Design of Experiments, and Survey Sampling*, Vol. 159 of Statistics: A Series of Textbooks and Monographs, New York: Dekker, pp. 401–432. [193]

Martin, E. C., and Betensky, R. A. (2005), "Testing Quasi-Independence of Failure and Truncation Times Via Conditional Kendall's Tau," *Journal of the American Statistical Association*, 100, 484–492. [192]

Mayer, T. (1975), "Selecting Economic Hypotheses by Goodness of Fit," *The Economic Journal*, 85, 877–883. [192]

McCann, R. J. (1995), "Existence and Uniqueness of Monotone Measure-Preserving Maps," *Duke Mathematics Journal*, 80, 309–323. [196]

Monge, G. (1781), "Mémoire sur la Théorie des Déblais et des Remblais," *Mémoires de l'Académie Royale des Sciences*, 1781, 666–704. [196]

Móri, T. F., and Székely, G. J. (2019), "Four Simple Axioms of Dependence Measures," *Metrika*, 82, 1–16. [199]

Mosteller, F. (1946), "On Some Useful "Inefficient" Statistics," *Annals Mathematical Statistics*, 17, 377–408. [200]

Möttönen, J., and Oja, H. (1995), "Multivariate Spatial Sign and Rank Methods," *Journal of Nonparameteric Statistics*, 5, 201–213. [202]

Mukaka, M. M. (2012), "A Guide to Appropriate Use of Correlation Coefficient in Medical Research," *Malawi Medical Journal*, 24, 69–71. [198]

Munkres, J. (1957), "Algorithms for the Assignment and Transportation Problems," *Journal of the Society for Industrial and Applied Mathematics*, 5, 32–38. [193,197]

Oja, H. (2010), *Multivariate Nonparametric Methods With R*, An Approach Based on Spatial Signs and Ranks, Vol. 199 of Lecture Notes in Statistics. New York: Springer. . [194,200]

Oja, H., and Randles, R. H. (2004), "Multivariate Nonparametric Tests," *Statistical Science*, 19, 598–605. [200]

Pearson, K. (1920), "Notes on the History of Correlation," *Biometrika*, 13, 25–45. [193,200]

Petrie, A. (2016), "Graph-Theoretic Multisample Tests of Equality in Distribution for High Dimensional Data," *Computational Statistics and Data Analysis*, 96, 145–158. [202]

Pillai, K. C. S., and K. Jayachandran (1967), "Power Comparisons of Tests of Two Multivariate Hypotheses Based on Four Criteria," *Biometrika*, 54, 195–210. [200]

Puri, M. L., and P. K. Sen (1966), "On a Class of Multivariate Multisample Rank-Order Tests," *Sankhyā*, Series A, 8, 353–376. [193]

——— (1971), *Nonparametric Methods in Multivariate Analysis*, New York: Wiley. [192,200]

Quessy, J.-F. (2010), "Applications and Asymptotic Power of Marginal-Free Tests of Stochastic Vectorial Independence," *Journal of Statistical Planning. Inference*, 140, 3058–3075. [200]

Randles, R. H., and Peters, D. (1990), "Multivariate Rank Tests for the Two-Sample Location Problem," *Communications in Statistics – Theory and Methods*, 19, 4225–4238. [202]

Rémillard, B. (2009), Discussion of "Brownian Distance Covariance," *Annals of Applied Statistics*, 3, 1295–1298. [201]

Rényi, A. (1959), "On Measures of Dependence," *Acta Mathematica Academiae Scientiarum Hungaricae*, 10, 441–451 (unbound insert). [200]

Reshef, D. N., Reshef, Y. A., Sabeti, P. C., and Mitzenmacher, M. (2018), "An Empirical Study of the Maximal and Total Information Coefficients and Leading Measures of Dependence," *Annals of Applied Statistics*, 12, 123–155. [200]

Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C., and Mitzenmacher, M. (2016), "Measuring Dependence Powerfully and Equitably," *Journal of Machine Learning Research*, 17, Paper No. 212, 63. [200]

Rizzo, M. L. (2009), "New Goodness-of-fit Tests for Pareto Distributions," *Astin Bulletin*, 39, 691–715. [195]

Rosenbaum, P. R. (2005), "An Exact Distribution-Free Test Comparing Two Multivariate Distributions Based on Adjacency," *Journal of Royal Statistical Society*, Series B, 67, 515–530. [193,195,202]

Rosenblatt, M. (1975), "A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence," *Annals of Statistics*, 3, 1–14. [200]

Rousson, V. (2002), "On Distribution-Free Tests for the Multivariate Two-Sample Location-Scale Model," *Journal of Multivariate Analysis*, 80, 43–57. [202]

Schilling, M. F. (1986), "Multivariate Two-Sample Tests Based on Nearest Neighbors," *Journal of the American Statistical Association*, 81, 799–806. [202]

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013), "Equivalence of Distance-Based and RKHS-Based Statistics in Hypothesis Testing," *Annals of Statistics*, 41, 2263–2291. [200,203]

Shi, H., Drton, M., and Han, F. (2020), "Distribution-Free Consistent Independence Tests Via Center-Outward Ranks and Signs," *Journal of the American Statistical Association*, 1–16. doi:10.1080/01621459.2020.1782223 [195,204]

Shi, H., Hallin, M., Drton, M., and Han, F. (2020), "Rate-Optimality of Consistent Distribution-Free Tests of Independence Based on Center-Outward Ranks and Signs," arXiv: 2007.02186. [204]

Smirnoff, N. (1939), "On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples," *Bulletin of Mathematical University Moscow*, 2, 16. [193]

Spearman, C. (1904), "The Proof and Measurement of Association Between Two Things," *American Journal of Psychology*, 15, 72–101. [200]

Székely, G. J., and Rizzo, M. L. (2005), "Hierarchical Clustering Via Joint Between-Within Distances: Extending Ward's Minimum Variance Method," *Journal of Classification*, 22, 151–183. [194,195]

Székely, G. J., and Rizzo, M. L. (2009), "Brownian Distance Covariance," *Annals of Applied Statistics*, 3, 1236–1265. [195,200,201]

Székely, G. J., and Rizzo, M. L. (2013), "Energy Statistics: A Class of Statistics Based on Distances," *Journal of Statistical Planning and Inference*, 143, 1249–1272. [194,195,199,203,204]

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *Annals of Statistics*, 35, 2769–2794. [194,198,199,200,201,202]

Taskinen, S., Kankainen, A., and Oja, H. (2003), "Sign Test of Independence Between Two Random Vectors," *Statistical & Probability Letters*, 62, 9–21. [200]

Taskinen, S., Oja, H., and Randles, R. H. (2005), "Multivariate Nonparametric Tests of Independence," *Journal of the American Statistical Association*, 100, 916–925. [194,200]

Villani, C. (2003), *Topics in Optimal Transportation*, Vol. 58 of Graduate Studies in Mathematics. Providence, RI: American Mathematical Society. [193,196]

Villani, C. (2009). *Optimal Transport*, Vol. 338 of *Grundlehren der Mathematischen Wissenschaften* [Fundamental Principles of Mathematical Sciences]. Berlin: Springer-Verlag. Old and new. [196]

Wald, A., and Wolfowitz, J. (1940), "On a Test Whether Two Samples Are From the Same Population," *Annals of Mathematics Statistics*, 11, 147–162. [193]

Weihs, L., Drton, M., and Meinshausen, N. (2018), "Symmetric Rank Covariances: A Generalized Framework for Nonparametric Measures of Dependence," *Biometrika*, 105, 547–562. [201]

Weiss, L. (1960), "Two-Sample Tests for Multivariate Distributions," *Annals of Mathematical Statistics*, 31, 159–164. [192]

Wilcoxon, F. (1947), "Probability Tables for Individual Comparisons by Ranking Methods," *Biometrics*, 3, 119–122. [193]

Wilks, S. S. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *The Annals of Mathematical Statistics*, 9, 60–62. [200]

Xiao, Y., Frisina, R., Gordon, A., Klebanov, L., and Yakovlev, A. (2004), "Multivariate Search for Differentially Expressed Gene Combinations," *BMC Bioinformatics*, 5, 164. [194]

Yang, G. (2012), "The Energy Goodness-of-Fit Test for Univariate Stable Distributions," Ph.D. thesis, Bowling Green State University. [195]

Zemel, Y., and Panaretos, V. M. (2019), "Fréchet Means and Procrustes Analysis in Wasserstein Space," *Bernoulli*, 25, 932–976. [197]

Zhao, J., and Meng, D. (2015), "FastMMD: Ensemble of Circular Discrepancy for Efficient Two-Sample Test," *Neural Computing*, 27, 1345–1372. [203]

Zuo, Y., and Serfling, R. (2000), "General Notions of Statistical Depth Function," *Annals of Statistics*, 28, 461–482. [193]