Distribution-free Measures of Association based on Optimal Transport

Nabarun Deb[1], Promit Ghosal[†2] and Bodhisattva Sen[*3]

[1] *University of Chicago, e-mail:* nabarun.deb@chicagobooth.edu

[2] *University of Chicago, e-mail:* promit@uchicago.edu

[3] *Columbia University, e-mail:* bodhi@stat.columbia.edu

In this paper we propose and study a class of nonparametric, yet interpretable measures of association between two random vectors X and Y taking values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ respectively ($d_1, d_2 \geq 1$). These nonparametric measures — defined using the theory of reproducing kernel Hilbert spaces coupled with optimal transport — capture the strength of dependence between X and Y and have the property that they are 0 if and only if the variables are independent and 1 if and only if one variable is a measurable function of the other. Further, these population measures can be consistently estimated using the general framework of geometric graphs which include k-nearest neighbor graphs and minimum spanning trees. Additionally, these measures can also be readily used to construct an exact finite sample distribution-free test of mutual independence between X and Y. In fact, as far as we are aware, these are the only procedures that possess all the above mentioned desirable properties. The correlation coefficient proposed in Dette et al. [25], Chatterjee [16] and Azadkia and Chatterjee [4], at the population level, can be seen as a special case of this general class of measures.

MSC2020 subject classifications: Primary 62G10, 62H20; secondary 60F05, 60D05.

Keywords and phrases: Maximum mean discrepancy, minimum spanning trees, multivariate ranks based on optimal transport, nearest neighbor graphs, reproducing kernel Hilbert spaces, testing for mutual independence, uniform central limit theorem.

1. Introduction

Suppose that $Z = (X, Y) \sim \mu$ where μ is an absolutely continuous distribution (with respect to the Lebesgue measure) supported on some subset of $\mathbb{R}^{d_1+d_2}$ ($d_1, d_2 \geq 1$) with marginal distributions μ_X and μ_Y on $\mathcal{X} \subset \mathbb{R}^{d_1}$ and $\mathcal{Y} \subset \mathbb{R}^{d_2}$ respectively. Assume that we have i.i.d. data $\{Z_i \equiv (X_i, Y_i)\}_{i=1}^n$ from μ. Our goal is to construct a real-valued measure to quantify the degree of association or dependence between X and Y, both in the population and sample settings.

Pearson's correlation coefficient (see e.g., [53]) is perhaps the simplest measure of association between X and Y when $d_1 = d_2 = 1$. However, the classical correlation has the following drawback: It is not effective for detecting associations that are nonlinear in nature, even in the complete absence of noise. Recently, Chatterjee [16] (also see [25]) proposed a *nonparametric* measure of association between X and Y when $d_1 = d_2 = 1$ which is 0 if and only if X and Y are independent and 1 if and only if Y is a measurable function of X. Moreover, any value between 0 and 1 of the coefficient conveys an idea of the strength of the relationship between

X and Y. In Azadkia and Chatterjee [4] the authors extend this measure to the case when $d_1 \geq 1$ and $d_2 = 1$. Deb et al. [22] propose a class of simple, nonparametric, yet interpretable measures of association between X and Y when they take values in general metric spaces. The above papers also propose and study consistent estimators of these population measures of dependence using ideas based on *geometric graphs* (e.g., k-nearest neighbor graph, minimum spanning tree, etc.).

To motivate our contribution in this paper consider the Spearman's correlation between X and Y (see [64]) which is another simple and useful measure of dependence when $d_1 = d_2 = 1$. An important property of the sample Spearman's correlation coefficient (not possessed by Pearson's correlation coefficient) is that it is *distribution-free*, i.e., its distribution does not depend on μ when X and Y are independent (also see Kendall's τ [43, 44]). This, in particular, implies that the sample Spearman's correlation coefficient can not only be used as a measure of dependence but can also be readily used to construct an *exact distribution-free procedure* to test the hypothesis of mutual independence between X and Y (when $d_1 = d_2 = 1$).

This leads to the following question: Can we construct a *nonparametric measure of association* between Y and X such that its sample analogue is *distribution-free* when $d_1, d_2 \geq 1$?

In this paper, we answer this question in the affirmative and propose and study a class of *distribution-free* empirical measures $T_n \equiv T_n(Z_1, \ldots, Z_n)$ and their population counterparts when $d_1, d_2 \geq 1$, that yield a family of *nonparametric measures of association* between Y and X. Our work extends the framework of Deb et al. [22]. In particular, these measures are defined using the theory of reproducing kernel Hilbert spaces (RKHS); see Section 2.1 for a brief review.

A plethora of nonparametric procedures have been proposed that can detect nonlinear dependencies between the variables X and Y over the last 60 years; see e.g., [6, 11, 12, 29, 31, 33, 37, 42, 46, 52, 54, 55, 56, 57, 67, 69] and the references therein. While these coefficients are useful in practice, they have one common problem: They are all designed primarily for testing independence, and not for measuring the strength of the relationship between the variables. Moreover, as far as we are aware, when $d_1, d_2 > 1$ none of above sample measures are distribution-free when X and Y are independent.

To construct our distribution-free measures of association we borrow ideas from three seemingly disparate fields: (i) The theory of RKHS which is needed to define the class of *nonparametric* measures of association; (ii) the notion of geometric graphs (see Section 2.3 for a review) which are used to develop *consistent* estimators of these nonparametric measures; and (iii) the theory of *optimal transport* (OT) which yields *distribution-free* empirical measures.

Our starting point is the class of kernel based measures of association defined in Deb et al. [22] (see Section 2.2 for a brief review). Although the proposed empirical measure there (denoted there by $\widehat{\eta}_n$) is nonparametric, it is not distribution-free. A key observation in this regard is that the distribution-free measures discussed above when $d_1 = d_2 = 1$ (e.g., Spearman's correlation coefficient) are based on the (univariate) "ranks" of Y_i's. Hence, one could imagine that a 'proper' distribution-free notion of multivariate ranks can be employed to construct distribution-free notions of dependence when $d_1, d_2 > 1$. Indeed this is our approach: We use the recently developed idea of *multivariate ranks* based on the theory of OT (see Sections 2.4 and 2.5 for a brief review on this topic; also see [19, 23, 59]) to develop measures of association that are finite-sample distribution-free when $\mu = \mu_X \otimes \mu_Y$.

Having defined the multivariate ranks (via OT) we construct our family of distribution-

free measures of association based on a very simple and classical analogy between Pearson's correlation and Spearman's correlation. Note that when $d_1 = d_2 = 1$, Spearman's correlation is equivalent to the classical Pearson's correlation coefficient computed between the one-dimensional ranks of the X_i's and the Y_i's, instead of using the observations themselves. We mimic the same approach here, i.e., instead of computing $\widehat{\eta}_n$ (as in [22]) using the X_i's and Y_i's themselves, we instead use their empirical multivariate ranks.

We propose the "rank" version of $\widehat{\eta}_n$, namely $\widehat{\eta}_n^{\text{RANK}}$, in (3.1). In Theorem 3.1, we show that $\widehat{\eta}_n^{\text{RANK}}$ consistently estimates a population measure of dependency η_K^{RANK} between X and Y, and is distribution-free when $\mu = \mu_X \otimes \mu_Y$. In Theorem 3.2 we show that this population measure η_K^{RANK} satisfies many desirable properties that justifies it as a population measure of association between X and Y. In particular, we show that:

(P1) $\eta_K^{\text{RANK}} \in [0, 1]$;

(P2) $\eta_K^{\text{RANK}} = 0$ if and only if $\mu = \mu_X \otimes \mu_Y$ (i.e., X and Y are independent);

(P3) $\eta_K^{\text{RANK}} = 1$ if and only if $Y = g(X)$, μ almost everywhere (a.e.), for some measurable function $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.

As $\widehat{\eta}_n^{\text{RANK}}$ is based on multivariate ranks, a test for independence of X and Y based on $\widehat{\eta}_n^{\text{RANK}}$ will generally be more powerful against heavy-tailed alternatives and more robust to outliers and contamination (see [23, 41, 52] for related discussions). Further, the corresponding test, being distribution-free, also avoids asymptotic approximations or permutation ideas for determining rejection thresholds.

In Proposition 3.1, we prove that the limit of $\widehat{\eta}_n^{\text{RANK}}$, i.e., η_K^{RANK} exactly coincides with the limit of the coefficient in Azadkia and Chatterjee [4] (denoted by $T_n(Y, Z)$ in their paper) when $d_2 = 1$ for an appropriate choice of a kernel. Note that, unlike $\widehat{\eta}_n^{\text{RANK}}$, the empirical measure in [4] does not have the finite sample distribution-free property.

Finally, Theorem 3.3 proves a central limit theorem for $\widehat{\eta}_n^{\text{RANK}}$ which is uniform over a large class of (geometric) graphs. We would like to point out that unlike X_i's and Y_i's, their multivariate ranks are no longer independent among themselves which makes the CLT challenging to prove.

The paper is organized as follows: In Section 2 we briefly review the RKHS framework, the notion of geometric graphs, and multivariate ranks via OT. Our distribution-free non-parametric measure of association is introduced in Section 3 along with the main results of this paper. We conclude with some remarks and open questions in Section 4. Section 5 gives the proofs of all the main results in this paper.

1.1. Related works

In Dette et al. [25], the authors use the term "measure of regression dependence" for the three properties mentioned above and show that it is possible to define such a measure satisfying (P1)-(P3) when $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. The same population measure was rediscovered in Chatterjee [16] where the author also proposed a tuning parameter-free estimator of the same measure that can be computed in near linear time. Since then, the estimator in [16] has attracted a lot of attention (see e.g., [14, 60]). Further, in Azadkia and Chatterjee [4], the authors propose a similar measure when $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \mathbb{R}^{d_2}$, $d_2 \geq 1$. However, all these measures crucially use the canonical ordering of $\mathbb{R}$ and hence do not extend to the multivariate setting (where $\mathcal{X} = \mathbb{R}^{d_1}$ and $\mathcal{Y} = \mathbb{R}^{d_2}$ with $d_1, d_2 \geq 2$). Some multivariate measures of association satisfying (P1)-(P3) have been proposed in [13, 62, 68], following similar copula-based ideas

as in [25]; however, to the best of our knowledge, neither of these papers provide a consistent empirical estimate of their proposed measures of association, nor do the associated tests for independence have the distribution-free property.

In a number of other papers, e.g., [3, 24, 45, 51, 60, 72], the authors analyze different properties of the estimator in Chatterjee [16]. In [2], the authors provide an extension of the measure in [4] to a vector of endogenous variables whereas [30] connects copula based measures of association to dimension reduction principles. In another line of work [66], the authors provide a recipe to modify a given measure of dependence to yield a measure of association satisfying (P1)-(P3) above. Finally, we refer the reader to [15, 17] for a survey of the recent papers in this topic.

2. Preliminaries

In this section we briefly introduce three topics: (i) The (nonparametric) class of kernel measures of association as defined in Deb et al. [22] using the theory of RKHS; (ii) the notion of geometric graphs which will be used to develop consistent estimators of these nonparametric population measures; and (iii) the theory of OT which will be used crucially to obtain distribution-free empirical measures.

Let $\mathcal{X}$ and $\mathcal{Y}$ be topological spaces. Let $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ with joint distribution μ and marginals $X \sim \mu_X$ and $Y \sim \mu_Y$. Although we will eventually take $\mathcal{X} = \mathbb{R}^{d_1}$ and $\mathcal{Y} = \mathbb{R}^{d_2}$, for much of what we review in this section we do not actually need to make this assumption. Let $\mathcal{P}(\mathcal{Y})$ and $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be the set of all Borel probability measures on $\mathcal{Y}$ and $\mathcal{X} \times \mathcal{Y}$ respectively. Suppose that μ admits a regular conditional distribution $\mu_{Y|x}$ — the conditional distribution of Y given $X = x$; existence of regular conditional distributions can be guaranteed under mild conditions (see [28] for a survey).

2.1. RKHS: Some background

In this subsection we formally define some concepts from the theory of RKHS that will be used repeatedly in this paper. We start with the basic definition of a RKHS.

Definition 2.1 (Reproducing kernel Hilbert space (RKHS))**.** *Let $\mathcal{H}$ be a Hilbert space of real-valued functions defined on $\mathcal{Y}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$. A function $K : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is called a* reproducing kernel *if the following two conditions hold:*

1. For all $y \in \mathcal{Y}$, $K(\cdot, y) \in \mathcal{H}$.

2. For all $y \in \mathcal{Y}$ and $f \in \mathcal{H}$, $\langle f, K(\cdot, y) \rangle_{\mathcal{H}} = f(y)$.

If $\mathcal{H}$ admits a reproducing kernel, then it is termed as a RKHS.

By the Moore-Aronszajn Theorem (see e.g., [5, Theorem 3]) a symmetric, nonnegative definite kernel function $K(\cdot, \cdot)$ on $\mathcal{Y} \times \mathcal{Y}$ can be identified uniquely with a unique RKHS of real-valued functions on $\mathcal{Y}$ for which $K(\cdot, \cdot)$ is the reproducing kernel. Let us denote this RKHS by $\mathcal{H}_K$. Let us also assume that $\mathcal{H}_K$ is separable (this can be ensured under mild conditions[1], see e.g., [65, Lemma 4.33]).

The map $y \mapsto K(\cdot, y)$ from $\mathcal{Y}$ to $\mathcal{H}_K$ is often called the *feature map*. Further, the repro-

---

[1]For example, if $\mathcal{Y}$ is a separable space and $K(\cdot, \cdot)$ is continuous.

ducing property as stated in Definition 2.1 implies that

$$\langle K(\cdot, y), K(\cdot, \widetilde{y})\rangle_{\mathcal{H}_K} = K(y, \widetilde{y}), \qquad \text{for all } y, \widetilde{y} \in \mathcal{Y}. \tag{2.1}$$

In the following we define three concepts that will be crucial in defining η_K — the kernel measure of dependence as introduced in Deb et al. [22]. Suppose that $Y \sim \mu_Y$ has a probability distribution on $\mathcal{Y}$ such that $\mathbb{E}[\sqrt{K(Y, Y)}] < \infty$.

Definition 2.2 (Mean embedding)**.** *Define the following class of probability measures on $\mathcal{Y}$:*

$$\mathcal{M}_K^\theta(\mathcal{Y}) := \left\{ \nu \in \mathcal{P}(\mathcal{Y}) : \int_{\mathcal{Y}} K^\theta(y, y) \, d\nu(y) < \infty \right\}, \qquad \text{for } \theta > 0.$$

Let $\mu_Y \in \mathcal{M}_K^{1/2}(\mathcal{Y})$. Then the (kernel) mean embedding *of μ_Y into $\mathcal{H}_K$ is given by $m_K(\mu_Y) \in \mathcal{H}_K$ such that*

$$\langle f, m_K(\mu_Y)\rangle_{\mathcal{H}_K} = \int_{\mathcal{Y}} f(y) \, d\mu_Y(y), \qquad \text{for all } f \in \mathcal{H}_K. \tag{2.2}$$

In fact, one can write $m_K(\mu_Y) = \int K(\cdot, y) \, d\mu_Y(y) = \mathbb{E}_{\mu_Y}[K(\cdot, Y)]$. It is well-defined as a consequence of the Riesz representation theorem, see [1, 63] (equivalently also by Bochner's theorem, see [26, 27]). The map $\mu_Y \mapsto m_K(\mu_Y)$ with domain $\mathcal{M}_K^{1/2}(\mathcal{Y})$ can be viewed as a natural extension of the map $y \mapsto K(\cdot, y)$ with domain $\mathcal{Y}$.

In a similar vein we can also define the (kernel) *mean embeddings* of conditional distributions. For $X = x$, the (kernel) *conditional mean embedding* of $\mu_{Y|x}$ is defined as an element of $\mathcal{H}_K$ in the same way as (2.2) (also see [49, Section 4.1.1]). In other words, we can also write $m_K(\mu_{Y|x}) = \int K(\cdot, y) d\mu_{Y|x}(y) = \mathbb{E}_{\mu_{Y|x}}[K(\cdot, Y)]$.

Definition 2.3 (Maximum mean discrepancy)**.** *The difference between two probability distributions Q_1 and Q_2 in $\mathcal{M}_K^1(\mathcal{Y})$ can then be conveniently measured by*

$$\mathrm{MMD}_K(Q_1, Q_2) := \|m_K(Q_1) - m_K(Q_2)\|_{\mathcal{H}_K}$$

(here $m_K(Q_i)$ is the mean element of Q_i, for $i = 1, 2$) which is called the maximum mean discrepancy (MMD) between Q_1 and Q_2 (see [34, Definition 10]). The following alternative representation of the squared MMD is also known (see e.g., [32, Lemma 6], or simply use (2.1)):

$$\mathrm{MMD}_K^2(Q_1, Q_2) = \mathbb{E}[K(S, S')] + \mathbb{E}[K(W, W')] - 2\mathbb{E}[K(S, W)], \tag{2.3}$$

where S, S', W, W' are independent, $S, S' \sim Q_1$ and $W, W' \sim Q_2$.

Definition 2.4 (Characteristic kernel)**.** *The kernel $K(\cdot, \cdot)$ is said to be characteristic if and only if the map $Q \mapsto m_K(Q)$ is one-to-one on the domain $\mathcal{M}_K^1(\mathcal{Y})$, i.e.,*

$$m_K(Q_1) = m_K(Q_2) \quad \Longrightarrow \quad Q_1 = Q_2, \quad \text{for all } Q_1, Q_2 \in \mathcal{M}_K^1(\mathcal{Y}).$$

Note that the last condition is equivalent to $\langle m_K(Q_1), f\rangle_{\mathcal{H}_K} = \langle m_K(Q_2), f\rangle_{\mathcal{H}_K}$ for all $f \in \mathcal{H}_K$, i.e., $\mathbb{E}_{S \sim Q_1}[f(S)] = \mathbb{E}_{W \sim Q_2}[f(W)]$, for all $f \in \mathcal{H}_K$. A characteristic kernel implicitly implies that the associated RKHS is rich enough.

2.2. A kernel measure of association

In this subsection we review the class of nonparametric kernel measures of association introduced in Deb et al. [22]. Recall that $\mu_{Y|x}$ is the regular conditional distribution of Y given $X = x$ which we assume exists for all $x \in \mathcal{X}$. Assume that a *kernel* $K(\cdot, \cdot)$ — a symmetric, nonnegative definite function on $\mathcal{Y} \times \mathcal{Y}$ — exists on $\mathcal{Y} \times \mathcal{Y}$ and let $\mathcal{H}_K$ denote the induced separable RKHS (see Definition 2.1). Let $\langle \cdot, \cdot \rangle_{\mathcal{H}_K} : \mathcal{H}_K \times \mathcal{H}_K \to \mathbb{R}$ and $\| \cdot \|_{\mathcal{H}_K}$ denote the inner product and induced norm on $\mathcal{H}_K$.

Generate $(X', Y', \widetilde{Y'})$ as follows: $X' \sim \mu_X$, $Y'|X \sim \mu_{Y|X'}$, $\widetilde{Y'}|X' \sim \mu_{Y|X'}$ and $Y' \perp\!\!\!\perp \widetilde{Y'}|X'$. Also let Y_1, Y_2 be i.i.d. from μ_Y. Note that Y' and $\widetilde{Y'}$ are dependent (via X'), unlike Y_1 and Y_2. The measure of association of Y on X as introduced in Deb et al. [22] can now be presented as:

$$\eta_K(\mu) := 1 - \frac{\mathbb{E}\|K(\cdot, Y') - K(\cdot, \widetilde{Y'})\|_{\mathcal{H}_K}^2}{\mathbb{E}\|K(\cdot, Y_1) - K(\cdot, Y_2)\|_{\mathcal{H}_K}^2}. \tag{2.4}$$

In order to ensure that $\eta_K(\cdot)$ is well-defined, we need certain moment assumptions. By the reproducing property of $K(\cdot, \cdot)$,

$$\|K(\cdot, Y_1) - K(\cdot, Y_2)\|_{\mathcal{H}_K}^2 = K(Y_1, Y_1) + K(Y_2, Y_2) - 2K(Y_1, Y_2). \tag{2.5}$$

Suppose that $\mu_Y \in \mathcal{M}_K^1(\mathcal{Y})$. Then the first two terms in (2.5) have finite moments. The third term is also finite by an application of the Cauchy-Schwartz inequality combined with the observation that $\mathcal{M}_K^1(\mathcal{Y}) \subset \mathcal{M}_K^{1/2}(\mathcal{Y})$. Thus, we assume that $\mu_Y \in \mathcal{M}_K^1(\mathcal{Y})$ in the sequel. In the following we give an alternate expression of η_K which will be crucial and can be easily derived from (2.4) and (2.5) (also see [22]):

$$\eta_K(\mu) = \frac{\mathbb{E}[K(Y_1, Y_2)] - \mathbb{E}[K(Y', \widetilde{Y'})]}{\mathbb{E}[K(Y_1, Y_1)] - \mathbb{E}[K(Y_1, Y_2)]}. \tag{2.6}$$

The following result shows that η_K is indeed a valid measure of association with many desirable properties.

Theorem 2.1 (Theorem 2.1 of [22]). *Suppose $\mu_Y \in \mathcal{M}_K^1(\mathcal{Y})$, $\mathcal{Y}$ is Hausdorff[2] and $K(\cdot, \cdot)$ is a characteristic kernel. Then $\eta_K(\mu)$, as defined in (2.4), satisfies the following properties:*

(P1) $\eta_K(\mu) \in [0, 1]$.

(P2) $\eta_K(\mu) = 0$ *if and only if $\mu = \mu_X \otimes \mu_Y$ (i.e., X and Y are independent).*

(P3) $\eta_K(\mu) = 1$ *if and only if $Y = g(X)$, μ almost everywhere (a.e.), for some measurable function $g : \mathcal{X} \to \mathcal{Y}$.*

2.3. Estimation via geometric graphs

In this subsection we review estimation of η_K (as in [22]) using ideas from geometric graphs. Note that the denominator in (2.6) can be easily estimated using empirical averages (from standard U-statistics theory; see [70, Chapter 12]), for instance, with the following estimator:

$$\frac{1}{n} \sum_{i=1}^n K(Y_i, Y_i) - \frac{1}{n(n-1)} \sum_{i \neq j} K(Y_i, Y_j).$$

---

[2]A topological space where for any two distinct points there exist neighborhoods of each which are disjoint from each other.

The numerator in (2.6) is trickier to estimate. The main difficulty arises because of the term $\mathbb{E}\big[K(Y',\widetilde{Y'})\big] = \mathbb{E}\big[\mathbb{E}(K(Y',\widetilde{Y'}) \mid X')\big]$ as we do not observe two Y values — Y' and $\widetilde{Y'}$ — independently from the conditional distribution of X' (cf. the notation introduced just before (2.4)). To overcome this difficulty, to estimate the above term we consider a 'neighbor' of X', say $\widetilde{X'}$, and its corresponding Y value, which can be used as a surrogate to $\widetilde{Y'}$. We formalize this notion of a 'neighbor' using a geometric graph on $\mathcal{X}$.

$\mathcal{G}$ is a *geometric graph functional* on $\mathcal{X}$ if, given any finite subset $S \subset \mathcal{X}$, $\mathcal{G}(S)$ defines a graph with vertex set S and corresponding edge set, say $\mathcal{E}(\mathcal{G}(S))$. Note that the graph $\mathcal{G}(S)$ can be *directed/undirected*. In this paper, we will restrict ourselves to simple graphs (i.e., those without multiple edges and self loops) with no isolated vertices. Accordingly, we will often drop the qualifier geometric and simple.

Next we define $\mathcal{G}_n := \mathcal{G}(X_1,\ldots,X_n)$ where $\mathcal{G}$ is some graph functional on $\mathcal{X}$. We would like to define graph functionals for which $(i,j) \in \mathcal{E}(\mathcal{G}_n)$ implies X_i and X_j are "close". Deb et al. [22] considered the following sample analogue of $\mathbb{E}\big[K(Y',\widetilde{Y'})\big]$:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_i} \sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)} K(Y_i, Y_j), \tag{2.7}$$

where $\mathcal{E}(\mathcal{G}_n)$ denotes the set of (directed/undirected) edges of $\mathcal{G}_n$, i.e., $(i,j) \in \mathcal{E}(\mathcal{G}_n)$ if and only if there is an edge from $i \to j$ or $j \to i$ in $\mathcal{G}_n$, and d_i denotes the degree of X_i in $\mathcal{G}_n$. To be specific $d_i := \sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)} 1$.

Using (2.7), Deb et al. [22] proposed the following estimator of η_K:

$$\widehat{\eta}_n := \frac{\frac{1}{n}\sum_{i=1}^{n} d_i^{-1} \sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)} K(Y_i, Y_j) - \frac{1}{n(n-1)}\sum_{i\neq j} K(Y_i, Y_j)}{\frac{1}{n}\sum_{i=1}^{n} K(Y_i, Y_i) - \frac{1}{n(n-1)}\sum_{i\neq j} K(Y_i, Y_j)}. \tag{2.8}$$

Under suitable conditions on the graph functional, [22] showed that indeed $\widehat{\eta}_n$ consistently estimate η_K as the sample size grows to infinity.

2.4. A brief overview of OT

In this subsection we introduce the basics of OT which will be necessary to construct distribution-free estimates of our measures of association. Let $\mathcal{P}_{ac}(\mathbb{R}^d)$ denote the space of absolutely continuous probability measures on $\mathbb{R}^d$. For a function $F : \mathbb{R}^d \to \mathbb{R}^d$, we will use $F\#\mu$ to denote the push forward measure of μ under F, i.e., the distribution of $F(Z)$ when $Z \sim \mu$.

Below we present perhaps the simplest version of the OT problem (courtesy the works of Gaspard Monge in 1781, see [48]):

$$\inf_F \int \|z - F(z)\|^2 \, d\mu(z) \qquad \text{subject to} \quad F\#\mu = \nu. \tag{2.9}$$

A minimizer of (2.9), if it exists, is referred to as an *OT map*. An important result in this field, known as the *Brenier-McCann theorem*, takes a "geometric" approach to the problem of OT (as opposed to the analytical approach presented in (2.9)) and will be very useful to us in the sequel; see e.g., [71, Theorem 2.12 and Corollary 2.30].

Proposition 2.1 (Brenier-McCann theorem [47])**.** *Suppose that $\mu, \nu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. Then there exists functions $R(\cdot)$ and $Q(\cdot)$ (usually referred to as "OT maps"), both of which are gradients of (extended) real-valued d-variate convex functions, such that: (i) $R\#\mu = \nu$, $Q\#\nu = \mu$; (ii) R and Q are unique (μ and ν-a.e. respectively); (iii) $R \circ Q(u) = u$ (ν-a.e. u) and $Q \circ R(z) = z$ (μ-a.e. z). Moreover, if μ and ν have finite second moments, $R(\cdot)$ is also the solution to Monge's problem in* (2.9).

In Proposition 2.1, by "gradient of a convex function" we essentially mean 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. It is instructive to note that, when $d = 1$, the standard 1-dimensional distribution function F associated with a distribution μ is nondecreasing and hence the gradient of a convex function. Therefore when $d = 1$, F is the OT map from μ to $\mathcal{U}[0, 1]$ (i.e., the Uniform([0,1]) distribution) by Proposition 2.1.

2.5. Multivariate ranks defined via OT

Definition 2.5 (Population multivariate ranks and quantiles)**.** *Set $\nu := \mathcal{U}[0, 1]^d$ — the uniform distribution on $[0, 1]^d$. Given $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$, the corresponding population rank and quantile maps are defined as the OT maps $R(\cdot)$ and $Q(\cdot)$ respectively as in Proposition 2.1. These maps are unique a.e. with respect to μ and ν respectively.*

In standard statistical applications, the population rank map is not available to the practitioner. In fact, the only accessible information about μ comes in the form of empirical observations $Z_1, Z_2, \ldots, Z_n \overset{i.i.d.}{\sim} \mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. In order to estimate the population rank map from these observations, let us denote

$$\mathcal{H}_n^d := \{h_1^d, \ldots, h_n^d\} \tag{2.10}$$

to be a set of n vectors in $[0, 1]^d$. We would like the points in $\mathcal{H}_n^d$ to be "uniform-like", i.e., their empirical distribution, $n^{-1} \sum_{i=1}^n \delta_{h_i^d}$ should converge weakly to ν. In practice, for $d = 1$, we may take $\mathcal{H}_n^d$ to be the usual $\{i/n\}_{1 \le i \le n}$ sequence and for $d \ge 2$, we may take it as a quasi-Monte Carlo sequence (such as the d-dimensional Halton or Sobol sequence) of size n (see [39, 40] for details) or a random draw of n i.i.d. random variables from ν. The empirical distribution on $\mathcal{H}_n^d$ will serve as a discrete approximation of ν. We are now in a position to define the empirical multivariate rank function which will proceed via a discrete analogue of problem (2.9).

Definition 2.6 (Empirical rank map)**.** *Let S_n denote the set of all $n!$ permutations of $\{1, 2, \ldots, n\}$. Consider the following optimization problem:*

$$\widehat{\sigma}_n := \underset{\sigma \in S_n}{\operatorname{argmin}} \sum_{i=1}^n \|X_i - h_{\sigma(i)}^d\|^2. \tag{2.11}$$

Note that $\widehat{\sigma}_n$ is a.s. uniquely defined (for each n) as $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. The empirical ranks are then defined as:

$$\widehat{R}_n(X_i) = h_{\widehat{\sigma}_n(i)}^d, \qquad \text{for } i = 1, \ldots, n. \tag{2.12}$$

The optimization problem in (2.11) is combinatorial in nature, but it can be solved exactly in polynomial time (with worst case complexity $\mathcal{O}(n^3)$) using the Hungarian algorithm (see [9,

50] for details). For a comprehensive list of faster (approximate) algorithms, see [59, Section 5]. Moreover, when $d = 1$, if we choose $\mathcal{H}_n^1 = \{i/n\}_{i=1}^n$, then the empirical ranks, i.e., $\widehat{R}_n(X_i)$'s are exactly equal to the usual one-dimensional ranks.

A crucial reason behind defining empirical ranks as in Definition 2.6 is that due to the exchangeability of the X_i's, the vector of ranks, i.e., $(\widehat{R}_n(X_1), \ldots, \widehat{R}_n(X_n))$ is uniformly distributed over the following set:

$$\{(h_{\sigma(1)}^d, \ldots, h_{\sigma(n)}^d) : \sigma \in S_n\}.$$

This lends a distribution-free property to the empirical ranks (see [23, Proposition 2.2] for a formal statement). In other words, the distribution of $(\widehat{R}_n(X_1), \ldots, \widehat{R}_n(X_n))$ is free of $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$. Moreover, the empirical ranks are maximal ancillary (see e.g., [35]).

3. A multivariate rank based measure of association

In this section we develop a new class of measures of association and their sample analogues which have the distribution-free property. We will restrict our attention to the case when $\mathcal{X} = \mathbb{R}^{d_1}$ and $\mathcal{Y} = \mathbb{R}^{d_2}$, for $d_1, d_2 \geq 1$.

3.1. A distribution-free measure of association and its properties

We will construct an empirical measure of association with the distribution-free property based on a simple and classical analogy between Pearson's and Spearman's correlation. Note that when $d_1 = d_2 = 1$, Spearman's correlation is equivalent to the classical Pearson's correlation coefficient computed between the one-dimensional ranks of the X_i's and the Y_i's, instead of the actual observations. It is this usage of one-dimensional ranks that lends Spearman's correlation the distribution-free property when $\mu = \mu_X \otimes \mu_Y$. We will mimic the same approach in this section, i.e., instead of computing $\widehat{\eta}_n$ (as defined in (2.8)) using the X_i's and Y_i's themselves, we will instead use their empirical multivariate ranks.

Let us briefly recall the setting. We have $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{i.i.d.}{\sim} \mu \in \mathcal{P}_{ac}(\mathbb{R}^{d_1+d_2})$ with marginals μ_X and μ_Y, such that $\mu_X \in \mathcal{P}_{ac}(\mathbb{R}^{d_1})$ and $\mu_Y \in \mathcal{P}_{ac}(\mathbb{R}^{d_2})$. Let $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$ be two sets of n "uniform-like" points in dimension d_1 and d_2 respectively. We can then use Definition 2.6 to define the empirical multivariate rank vectors $(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n))$ and $(\widehat{R}_n^Y(Y_1), \ldots, \widehat{R}_n^Y(Y_n))$ based on $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$ (see (2.10)) respectively.

Next, given a graph functional $\mathcal{G}$ (see Section 2.3 for its definition), let

$$\mathcal{G}_n^{\text{RANK}} := \mathcal{G}(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n))$$

and $\mathcal{E}(\mathcal{G}_n^{\text{RANK}})$ be the set of edges of $\mathcal{G}_n^{\text{RANK}}$. Also, with a slight notational abuse, we will still use $(d_1, \ldots, d_n)$ to be the degree sequence of the vertices in $\mathcal{G}_n^{\text{RANK}}$. The rank version of $\widehat{\eta}_n$ is then defined as follows:

$$\widehat{\eta}_n^{\text{RANK}} := \frac{n^{-1} \sum_i d_i^{-1} \sum_{j:(i,j) \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) - F_n}{n^{-1} \sum_{i=1}^n K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_i)) - F_n}, \tag{3.1}$$

where

$$F_n := (n(n-1))^{-1} \sum_{i \neq j} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)). \tag{3.2}$$

Here $(i,j) \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})$ implies that $\widehat{R}_n^Y(X_i)$ and $\widehat{R}_n^Y(X_j)$ are connected in $\mathcal{G}_n^{\text{RANK}}$. Note that $\widehat{\eta}_n^{\text{RANK}}$ is the same as $\widehat{\eta}_n$ with the observations replaced with their empirical ranks. In the following theorems (proved in Section 5), we show that $\widehat{\eta}_n^{\text{RANK}}$ is a measure of association which additionally has a pivotal distribution when $\mu = \mu_X \otimes \mu_Y$.

Theorem 3.1. (a) *When $\mu = \mu_X \otimes \mu_Y$, $\widehat{\eta}_n^{\text{RANK}}$ has a pivotal distribution, i.e., its distribution does not depend on μ_X and μ_Y.*

(b) *Assume that $K(\cdot,\cdot) : \mathbb{R}^{d_2} \times \mathbb{R}^{d_2} \to \mathbb{R}$ is continuous and let $R^X(\cdot)$ and $R^Y(\cdot)$ denote the population rank maps (see Definition 2.5) for μ_X and μ_Y respectively. Suppose that the following assumptions hold:*

(S1) For $x_1, x_2 \in [0,1]^{d_1}$, define

$$r(x_1, x_2) := \mathbb{E}\Big[K(R^Y(Y_1), R^Y(Y_2)) \mid R^X(X_1) = x_1, R^X(X_2) = x_2 \Big]$$

and assume that it is uniformly β-Hölder continuous in $x_1, x_2 \in [0,1]^{d_1}$ for some $\beta \in (0,1]$, i.e., given any $x_1, x_2, \widetilde{x}_1, \widetilde{x}_2 \in [0,1]^{d_1}$, there exists a constant C (free of $x_1, x_2, \widetilde{x}_1, \widetilde{x}_2$) such that:

$$\big| r(x_1, x_2) - r(\widetilde{x}_1, \widetilde{x}_2) \big| \leq C \left(\|x_1 - \widetilde{x}_1\|^\beta + \|x_2 - \widetilde{x}_2\|^\beta \right). \tag{3.3}$$

(S2) Let $\mathcal{G}_n^{\text{RANK}}$ satisfy the following conditions:

$$\limsup_{n \to \infty} \frac{\max_{i=1}^n d_i}{\min_{i=1}^n d_i} < \infty, \tag{3.4}$$

$$\frac{1}{n \min_{i=1}^n d_i} \sum_{e \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})} |e|^\beta \to 0, \tag{3.5}$$

where $|e|$ denotes the edge length of $e \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})$.

(S3) The empirical distributions on $\mathcal{H}_n^{d_1}$ and $\mathcal{H}_n^{d_2}$ converge weakly to $\mathcal{U}[0,1]^{d_1}$ and $\mathcal{U}[0,1]^{d_2}$ respectively.

Under these assumptions,

$$\widehat{\eta}_n^{\text{RANK}} \xrightarrow{\mathbb{P}} 1 - \frac{\mathbb{E}\|K(\cdot, R^Y(Y')) - K(\cdot, R^Y(\widetilde{Y'}))\|_{\mathcal{H}_K}^2}{\mathbb{E}\|K(\cdot, R^Y(Y_1)) - K(\cdot, R^Y(Y_2))\|_{\mathcal{H}_K}^2} := \eta_K^{\text{RANK}},$$

where Y_1, Y_2 are i.i.d. from μ_Y, and $(Y', \widetilde{Y'})$ are defined as follows: Generate $(X', Y', \widetilde{Y'})$ as $X' \sim \mu_X$, $Y'|X \sim \mu_{Y|X'}$, $\widetilde{Y'}|X \sim \mu_{Y|X'}$ and $Y' \perp\!\!\!\perp \widetilde{Y'}|X'$.[3]

Remark 3.1 (On condition (S1)). *To better understand (3.3) let us first give an alternate expression for*

$$r(x_1, x_2) = \langle g(x_1), g(x_2) \rangle_{\mathcal{H}_K}, \quad \text{where} \quad g(x) := \mathbb{E}\big[K(R^Y(Y), \cdot) \mid R^X(x) = x \big],$$

i.e., $r(\cdot, \cdot)$ is the inner product between the conditional feature maps (after the rank transformations). Thus, (3.3) can be thought of as a Lipschitz assumption on the inner product. A

---

[3]Note that Y' and $\widetilde{Y'}$ are dependent (via X'), unlike Y_1 and Y_2.

similar condition also appears in [4, 22]. This condition is needed in our proof technique but we expect Theorem 3.1 to hold in more generality.

Let us illustrate how under reasonably practical model assumptions condition (S1) is satisfied. Suppose that one assumes a model where $Y = f(X) + \epsilon$, and f is a Borel measurable function, with the noise variable ϵ being independent of X. Additionally, suppose that $(R^X)^{-1}$ (inverse function of R^X), R^Y, f, and both arguments of the kernel $K(\cdot, \cdot)$ in Theorem 3.1 are β-Hölder continuous.[4] Then we can write

$$\left| r(x_1, x_2) - r(\widetilde{x}_1, \widetilde{x}_2) \right| \leq \mathbb{E}\Big[\big| K\big(R^Y(f((R^X)^{-1}(x_1)) + \epsilon_1), f((R^X)^{-1}(x_2)) + \epsilon_2)\big)$$
$$- K\big(R^Y(f((R^X)^{-1}(\widetilde{x}_1)) + \epsilon_1), R^Y(f((R^X)^{-1}(\widetilde{x}_2)) + \epsilon_2))\big| \Big]$$
$$\leq C(\|x_1 - \widetilde{x}_1\|^\beta + \|x_2 - \widetilde{x}_2\|^\beta)$$

where C is a constant that depends on the kernel K and the distributions of X and Y. The last inequality above follows from the β-Hölder continuity of $(R^X)^{-1}, R^Y, f$ and K.

Remark 3.2 (On condition (S2)). *As the set $\mathcal{H}_n^{d_1}$, the set of empirical multivariate ranks, i.e., $(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n))$ is some permutation of $\mathcal{H}_n^{d_1}$, note that $\max_{1 \leq i \leq n} d_i$, $\min_{1 \leq i \leq n} d_i$, and $\frac{1}{n \min_{i=1}^n d_i} \sum_{e \in \mathcal{E}(\mathcal{G}_n^{\mathrm{RANK}})} |e|^\beta$ are all deterministic quantities appearing in condition (S2) above. This should not be confused with the randomness inherent in the construction of $\mathcal{G}_n^{\mathrm{RANK}}$. For example, $\mathbf{1}((i, j) \in \mathcal{E}(\mathcal{G}_n^{\mathrm{RANK}}))$, i.e., indicator of the event that $\widehat{R}_n^X(X_i)$ and $\widehat{R}_n^X(X_j)$ are connected in $\mathcal{G}_n^{\mathrm{RANK}}$, is still a random variable.*

Theorem 3.1 (proved in Section 5) shows that $\widehat{\eta}_n^{\mathrm{RANK}}$ can be used to construct a test for independence (in addition to being a measure of association) which will be consistent and exactly distribution-free under the null hypothesis of independence.

The above result shows that not only is our measure of association $\widehat{\eta}_n^{\mathrm{RANK}}$ distribution-free when X and Y are independent, it converges to a population limit η_K^{RANK} as n grows large, under suitable assumptions. We call η_K^{RANK} as the *population measure of association*. Indeed, as the following result shows, η_K^{RANK} satisfies many desirable properties that justifies it as a population measure of association between X and Y.

Theorem 3.2. *If $K(\cdot, \cdot)$ is characteristic and continuous, then η_K^{RANK} satisfies the following properties:*

(P1) $\eta_K^{\mathrm{RANK}} \in [0, 1]$.

(P2) $\eta_K^{\mathrm{RANK}} = 0$ if and only if $\mu = \mu_X \otimes \mu_Y$ (i.e., X and Y are independent).

(P3) $\eta_K^{\mathrm{RANK}} = 1$ if and only if $Y = g(X)$, μ almost everywhere (a.e.), for some measurable function $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.

Remark 3.3 (Asymmetric nature of $\widehat{\eta}_n^{\mathrm{RANK}}$ and η_K^{RANK}). *Note that $\widehat{\eta}_n^{\mathrm{RANK}}$, as defined in (3.1), is not symmetric in the X_is and Y_js. This is by design: The goal of $\widehat{\eta}_n^{\mathrm{RANK}}$ is to measure how well Y can be predicted from X, which can be very different from how well X can be predicted from Y. For example, if $Y = f(X) + \epsilon$ for ϵ independent of X, $X \in \mathbb{R}^{d_1}$, $d_1 > 1$, and $f : \mathbb{R}^{d_1} \to \mathbb{R}$, then it is conceivable that it is easier to predict Y from X than to predict*

---

[4]The stated assumptions on R^X and R^Y are valid when the distributions of X and Y are supported on closed bounded sets, and their marginal density functions are bounded above and below in their support sets and are β-Hölder continuous [20, Corollary 5].

X from Y. Hence it is natural to construct $\widehat{\eta}_n^{\text{RANK}}$ in an asymmetric fashion. Of course, it is possible to symmetrize $\widehat{\eta}_n^{\text{RANK}}$ by switching the role of X, Y, and then taking a maximum. It is clear from Theorems 3.1 and 3.2 that this symmetrized version will also converge to a measure of association satisfying (P1), (P2), and a symmetric variant of (P3).

Next we show that η_K^{RANK} (which arises naturally as the limit of $\widehat{\eta}_n^{\text{RANK}}$ in Theorem 3.1) is exactly the same as the following population correlation coefficient proposed in Chatterjee [16, Equation 1.2] and Azadkia and Chatterjee [4, Equation 2.1] where the authors only consider the case $d_2 = 1$ (also see Dette et al. [25]):

$$\xi(\mu) := \frac{\int \text{Var}(\mathbb{P}(Y \geq t|X)) \, d\mu_Y(t)}{\int \text{Var}(\mathbf{1}(Y \geq t)) \, d\mu_Y(t)}$$

for a particular choice of a characteristic kernel. The choice turns out to be the kernel:

$$K(y_1, y_2) = |y_1| + |y_2| - |y_1 - y_2|. \tag{3.6}$$

Incidentally this is also the kernel used in the construction of distance covariance (see [67]). We state the result formally in the proposition below.

Proposition 3.1. *When $d_2 = 1$ and $K(\cdot, \cdot)$ is the kernel in (3.6), then $\eta_K^{\text{RANK}} = \xi(\mu)$.*

The above proposition shows that the family of measures of association proposed in this paper includes that of Chatterjee [16] and extends it significantly to general $d_1, d_2 \geq 1$ and also a large class of kernel functions.

3.2. A rank CLT under null

We present a CLT for $\widehat{\eta}_n^{\text{RANK}}$ when $\mu = \mu_X \otimes \mu_Y$. Note that $\widehat{\eta}_n^{\text{RANK}}$ is a function of $\widehat{R}_n(Y_i)$'s (respectively $\widehat{R}_n(X_i)$'s) which are no longer independent among themselves. As a result, proving a CLT for $\widehat{\eta}_n^{\text{RANK}}$ is not trivial. Informally speaking, the technique used in this paper is that of a Hájek representation (as in [60]), where we show that the empirical multivariate ranks can be replaced by their population counterparts in $\widehat{\eta}_n^{\text{RANK}}$ at a $o_{\mathbb{P}}(1/\sqrt{n})$ cost. As $\widehat{\eta}_n^{\text{RANK}}$ does not have a standard U-statistic representation, we do not use the explicit form of Hájek projections as in [60], but opt for a more hands-on method of moments based approach.

Note that

$$\sqrt{n} \, \widehat{\eta}_n^{\text{RANK}} = \frac{N_n^{\text{RANK}}}{D_n} \tag{3.7}$$

where

$$D_n := \frac{1}{n} \sum_{i=1}^{n} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_i)) - F_n$$

and

$$N_n^{\text{RANK}} := \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_i}^{-1} \sum_{j:(i,j) \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})} K\left(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)\right) - F_n \right) \tag{3.8}$$

with F_n defined in (3.2). Now, D_n does not involve the X_i's and converges to $\mathbb{E}[K(R^Y(Y_1), R^Y(Y_1))] - \mathbb{E}[K(R^Y(Y_1), R^Y(Y_2))]$ by the continuous mapping theorem. Therefore, by Slutsky's theorem, it suffices to establish a pivotal limit distribution for N_n^{RANK} after a suitable scaling. This is the subject of the following theorem.

Theorem 3.3. *If* (3.4) *holds,* $K(\cdot, \cdot)$ *is continuous,* $\mu = \mu_X \otimes \mu_Y$, $\mu_X \in \mathcal{P}_{ac}(\mathbb{R}^{d_1})$ *and* $\mu_Y \in \mathcal{P}_{ac}(\mathbb{R}^{d_2})$, *then* $\mathrm{Var}(N_n^{\mathrm{RANK}}) = \mathcal{O}(1)$. *Further, with* $\theta := (D, \gamma, \epsilon) \in (0, \infty)^3$ *consider the following subclass of graph functionals and measures on* $\mathcal{X} \times \mathcal{Y}$ *given by:*

$$\mathcal{J}_\theta := \left\{ \widetilde{\mathcal{G}} : \limsup_{n \to \infty} \left(\max_{1 \le i \le n} \frac{\widetilde{d}_i}{(\log n)^\gamma} + \frac{\max_{i=1}^n \widetilde{d}_i}{\min_{i=1}^n \widetilde{d}_i} \right) \le D, \ \mathrm{Var}(N_n^{\mathrm{RANK}}) \ge \epsilon \ \forall n \ge D, \right.$$

where $(\widetilde{d}_1, \dots, \widetilde{d}_n)$ *denotes the degree sequence of* $\widetilde{\mathcal{G}}_n^{\mathrm{RANK}} := \widetilde{\mathcal{G}}(\widehat{R}_n^X(X_1), \dots, \widehat{R}_n^X(X_n))\Big\}.$

Then, under assumption (S3) from Theorem 3.1, the following result holds for every fixed $\theta \in (0, \infty)^3$:

$$\lim_{n \to \infty} \sup_{\widetilde{\mathcal{G}} \in \mathcal{J}_\theta} \sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\frac{N_n^{\mathrm{RANK}}}{\widetilde{S}_n} \le z \right) - \Phi(z) \right| = 0, \tag{3.9}$$

where

$$\widetilde{S}_n^2 := \widetilde{a}\left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1} \right) + \widetilde{b}\left(\widetilde{g}_2 - 2\widetilde{g}_1 - 2\widetilde{g}_3 - \frac{n-5}{n-1} \right) + \widetilde{c}\left(\widetilde{g}_1 + \widetilde{g}_3 - \widetilde{g}_2 + \frac{n-3}{n-1} \right)$$

with

$$\widetilde{a} := \mathbb{E}\left[K^2\big(R^Y(Y_1), R^Y(Y_2)\big) \right],$$
$$\widetilde{b} := \mathbb{E}\left[K\big(R^Y(Y_1), R^Y(Y_2)\big) K\big(R^Y(Y_1), R^Y(Y_3)\big) \right],$$
$$\widetilde{c} := \left(\mathbb{E}\left[K\big(R^Y(Y_1), R^Y(Y_2)\big) \right] \right)^2,$$

and

$$\widetilde{g}_1 := \frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{d}_i}, \quad \widetilde{g}_2 := \frac{1}{n} \sum_{i,j} \frac{T^{\widetilde{\mathcal{G}}_n^{\mathrm{RANK}}}(i,j)}{\widetilde{d}_i \widetilde{d}_j}, \quad \widetilde{g}_3 := \frac{1}{n} \sum_{(i,j) \in \mathcal{E}(\widetilde{\mathcal{G}}_n^{\mathrm{RANK}})} \frac{1}{\widetilde{d}_i \widetilde{d}_j}.$$

Here

$$T^{\widetilde{\mathcal{G}}_n^{\mathrm{RANK}}}(i,j) := \sum_k \mathbf{1}\big((i,k) \in \mathcal{E}(\widetilde{\mathcal{G}}_n^{\mathrm{RANK}})\big) \mathbf{1}\big((k,j) \in \mathcal{E}(\widetilde{\mathcal{G}}_n^{\mathrm{RANK}})\big)$$

denotes the number of common neighbors of $\widehat{R}_n^X(X_i)$ *and* $\widehat{R}_n^X(X_j)$.

Remark 3.4 ($\widetilde{S}_n^2$ *is distribution-free*)**.** *It is worth noting that although* $\widetilde{S}_n^2$ *has a complicated expression, it does not depend on the underlying distributions* μ_X *and* μ_Y. *In particular,* $\widetilde{a}, \widetilde{b}, \widetilde{c}$ *are fixed constants as* $R^Y(Y_1), R^Y(Y_2), R^Y(Y_3) \overset{i.i.d.}{\sim} \mathcal{U}[0,1]^{d_2}$. *Moreover,* $\widetilde{g}_1, \widetilde{g}_2, \widetilde{g}_3$ *are deterministic quantities based on a graph computed from a fixed set of points.*

Remark 3.5 (*Uniformity over* $\widetilde{\mathcal{G}}$ *in Theorem 3.3*)**.** *One of the strengths of Theorem 3.3 is that it is uniform over a large class of graphs (see* (3.9)). *This implies that the convergence to normality will hold for data dependent choices of graphs; e.g., when using a k-nearest neighbor graph with k chosen as a function of the observed data, provided the choice only grows logarithmically in n with probability converging to 1 (see the definition of* $\mathcal{J}_\theta$).

4. Discussion and open questions

To summarize, in the current manuscript, we have provided a measure of association between two random variables $X \in \mathcal{X} \subset \mathbb{R}^{d_1}$ and $Y \in \mathcal{Y} \subset \mathbb{R}^{d_2}$ which equals 0 or 1 depending on

whether Y is independent of X or Y is a noiseless measurable function of X. We propose a class of tuning-free estimators of this measure which has the added benefit of being exactly distribution-free under the null hypothesis of independence between X and Y. Further, we prove a uniform central limit theorem under the null. At a broader conceptual level, our work brings together three different disciplines — RKHSs, geometric graphs, and OT, in order to achieve the aforementioned set of properties. Our work opens up a number of potential directions to explore in the future. We outline some of them below.

(a) **CLT under alternative**: When $\mu \neq \mu_X \otimes \mu_Y$, the question of proving a limit theorem for $\widehat{\eta}_m^{\mathrm{RANK}}$ or N_n^{RANK} remains open. The main source of difficulty lies in the fact that N_n^{RANK} no longer has mean 0 and in fact, the resulting bias decays slower than $n^{-1/2}$. This suggests the need of debiasing $\widehat{\eta}_n^{\mathrm{RANK}}$ appropriately. We believe that the works of [7, 8] could be relevant in this direction.

(b) **Monotonicity**: As $\eta_K^{\mathrm{RANK}} = 0$ and 1 capture independence and perfect dependence between Y and X, it is natural to expect that the relationship between Y and X should get stronger continuously and monotonically as η_K^{RANK} increases from 0 to 1. For a variant of η_K^{RANK} in the univariate $d_1 = d_2 = 1$ setting, this monotonicity was demonstrated for certain natural examples in [3, Proposition 4.1]. It would be interesting to establish such monotonicity for general kernels and for $d_1, d_2 > 1$.

(c) **Local power analysis**: The question of finding the detection boundary for testing independence using N_n^{RANK} is of considerable interest. To put it informally, the questions we ask here is: What is the minimum "distance" between μ and $\mu_X \otimes \mu_Y$, such that the test for independence using $\widehat{\eta}_n^{\mathrm{RANK}}$ has asymptotic power 1? Based on the works of [3, 45, 60], it is evident that the answer should depend heavily on the choice of the graph functional used in the construction of $\widehat{\eta}_n^{\mathrm{RANK}}$, e.g., the choice of k in a k-nearest neighbor graph. As in (a) above, the hardness arises again in tightly quantifying the bias in N_n^{RANK}. Here we expect a "blessing of dimensionality" in the detection boundary beyond $d_2 = 8$ (see Bhattacharya [10] for a related result).

(d) **Numerical performance**: We do not pursue a thorough simulation study comparing the power of the distribution-free test of independence based on $\widehat{\eta}_n^{\mathrm{RANK}}$ with other competing methods here. However, based on recent experimental observations made for other graph-based and rank-based tests, there are a number of interesting practical questions one may ask. For instance, graph-based tests usually outperform competitors when the relationship between Y and X is sinusoidal (see [6, 16]) while multivariate rank-based tests are known for their robustness against outliers and heavy-tailed data (see [23, 61]). As $\widehat{\eta}_n^{\mathrm{RANK}}$ uses both, one may ask: Does it inherit the best of both worlds? Also, in [21, Section D.5], the authors show that certain multivariate rank-based tests have competitive performance in high-dimensional problems as well. It would be interesting to explore if $\widehat{\eta}_n^{\mathrm{RANK}}$ exhibits similar behavior.

(e) **Extension beyond uniform reference distribution**: Throughout this paper we have assumed that the reference distributions needed for defining the multivariate ranks are uniform on the d_1 and d_2-dimensional unit hypercube. This choice is made mostly for simplicity and from a methodological perspective any absolutely continuous reference distribution would have sufficed (see e.g., [21, 36, 58] for some alternate choices of the reference distribution). Our proofs would work verbatim for any other compactly supported reference distribution. However, for reference distributions with unbounded supports (e.g., multivariate normal) some truncation arguments would be needed. The benefits of using other reference distributions have a rich literature dating back to [18, 38] and more recently in [21, 58]. It re-

mains open to explore whether such gains can be had by choosing other reference distributions for constructing $\widehat{\eta}_n^{\text{RANK}}$.

5. Proofs of our main results

5.1. Proof of Theorem 3.1

First we state an important result about the consistency of the empirical multivariate ranks (as in Definition 2.6), i.e., $\widehat{R}_n(\cdot)$ yields a consistent estimate of $R(\cdot)$ (see [23] for a proof).

Proposition 5.1. *Suppose $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and $n^{-1} \sum_{i=1}^n \delta_{h_i^d} \xrightarrow{w} \nu$. Then $n^{-1} \sum_{i=1}^n \|\widehat{R}_n(X_i) - R(X_i)\|_2 \xrightarrow{a.s.} 0$.*

Proof of (i): By [23, Proposition 2.2 (ii)], $(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n))$ (or $(\widehat{R}_n^Y(Y_1), \ldots, \widehat{R}_n^Y(Y_n))$) is distributed uniformly over all $n!$ permutations of the set $\mathcal{H}_n^{d_1}$ (or $\mathcal{H}_n^{d_2}$). When $\mu = \mu_X \otimes \mu_Y$, clearly $(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n))$ and $(\widehat{R}_n^Y(Y_1), \ldots, \widehat{R}_n^Y(Y_n))$ are independent and the joint distribution is pivotal. As $\widehat{\eta}_n^{\text{RANK}}$ is a function of $(\widehat{R}_n^X(X_1), \ldots, \widehat{R}_n^X(X_n), \widehat{R}_n^Y(Y_1), \ldots, \widehat{R}_n^Y(Y_n))$, it also has a pivotal distribution.

Proof of (ii): As $n^{-1} \sum_{i=1}^n \delta_{h_i^{d_2}}$ converges weakly to $\mathcal{U}[0, 1]^{d_2}$, the following conclusions are easy consequences of the Portmanteau Theorem:

$$\frac{1}{n} \sum_{i=1}^n K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_i)) \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}K(R^Y(Y_1), R^Y(Y_1)),$$

$$\frac{1}{n(n-1)} \sum_{i \neq j} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}K(R^Y(Y_1), R^Y(Y_2)). \tag{5.1}$$

In the above displays, both terms on the right hand side are deterministic because both of them are permutation-invariant and $(\widehat{R}_n(Y_1), \ldots, \widehat{R}_n(Y_n))$ is some permutation of the fixed set $\mathcal{H}_n^{d_2}$. Consequently the above convergence is a deterministic result. By an application of Proposition 5.1 and (3.4), it further suffices to show that:

$$\mathcal{Z}_n := \frac{1}{n} \sum_i d_i^{-1} \sum_{j : (i,j) \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})} K(R^Y(Y_i), R^Y(Y_j)) \stackrel{\mathbb{P}}{\longrightarrow} \mathbb{E}K(R^Y(Y'), R^Y(\widetilde{Y'})).$$

In order to establish the above, by an application of Chebyshev's inequality, it suffices to show that:

(a) $\mathbb{E}\mathcal{Z}_n^2 \stackrel{n \to \infty}{\longrightarrow} \left[\mathbb{E}K(R^Y(Y'), R^Y(\widetilde{Y'})) \right]^2$.

(b) $\mathbb{E}\mathcal{Z}_n \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}K(R^Y(Y'), R^Y(\widetilde{Y'}))$.

We first prove (a). For the sake of simplicity, we write

$$\mathbb{E}\mathcal{Z}_n^2 = (\mathbf{I}) + (\mathbf{II}) + (\mathbf{III})$$

where

$$(\mathbf{I}) := \frac{1}{n^2} \mathbb{E}\left[\sum_i \sum_{j : (i,j) \in \mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \frac{1}{d_i} \left(\frac{1}{d_i} + \frac{1}{d_j} \right) \mathbb{E}\left[K^2(R^Y(Y_i), R^Y(Y_j)) \mid R^X(X_i), R^X(X_j) \right] \right]$$

$$(\textbf{II}) := \frac{1}{n^2}\mathbb{E}\Bigg[\sum_i \sum_{j \neq k:(i,j),(i,k)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \left(\frac{1}{d_i}+\frac{1}{d_j}\right)\left(\frac{1}{d_i}+\frac{1}{d_k}\right)$$

$$\times \mathbb{E}\Big[K(R^Y(Y_i),R^Y(Y_j))\,K(R^Y(Y_i),R^Y(Y_k)) \mid R^X(X_i),R^X(X_j),R^X(X_k)\Big]\Bigg]$$

$$(\textbf{III}) := \frac{1}{n^2}\mathbb{E}\Bigg[\sum_{i\neq j} \sum_{k\neq\ell:(k,i),(\ell,j)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \frac{1}{d_i d_j}\mathbb{E}\Big[K(R^Y(Y_i),R^Y(Y_k))$$

$$\times K(R^Y(Y_j),R^Y(Y_\ell)) \mid R^X(X_i),R^X(X_j),R^X(X_k),R^X(X_l)\Big]\Bigg].$$

We now claim and prove that

$$(\textbf{III}) \to \left(\mathbb{E}\Big(\mathbb{E}\Big[K\big(R^Y(\widetilde{Y}_1),R^Y(\widetilde{Y}_2)\big) \mid X\Big]\Big)\right)^2 \tag{5.2}$$

as $n\to\infty$, where $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are independent samples from the conditional distribution $Y\mid X$. For this, we define

$$(\widetilde{\textbf{III}}) := \frac{1}{n^2}\sum_{i\neq j}\sum_{k\neq\ell:(k,i),(\ell,j)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \frac{1}{d_i}\frac{1}{d_j}\mathbb{E}\Big[\mathbb{E}\Big[K\big(R^Y(Y_i),R^Y(Y_i')\big) \mid R^X(X_i)\Big]$$

$$\times \mathbb{E}\Big[K\big(R^Y(Y_j),R^Y(Y_j')\big) \mid R^X(X_j)\Big]\Big]$$

where Y_i' (resp. Y_j') is a sample from the conditional distribution $Y\mid X_i$ (resp. $Y\mid X_j$) independent of Y_i and Y_j. A simple application of the fact that $K(\cdot,\cdot)$ is continuous a.e., $R^Y(\cdot)$ is bounded coordinate-wise, shows that:

$$|(\textbf{III}) - (\widetilde{\textbf{III}})|$$
$$\lesssim C\frac{\max_{j=1}^n d_j}{n\min_{j=1}^n d_j}\sum_i \frac{1}{d_i}\mathbb{E}\Big[\sum_{k:(k,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \big|\mathbb{E}\big[K(R^Y(Y_i),R^Y(Y_k)) \mid R^X(X_i),R^X(X_k)\big]$$
$$- \mathbb{E}\big[K(R^Y(Y_i),R^Y(Y_i')) \mid R^X(X_i)\big]\big|\Big]. \tag{5.3}$$

We show that the right hand side of the above inequality converges to 0 as $n\to\infty$. By the assumption (3.3), $\mathbb{E}\big[K\big(R^Y(Y_1),R^Y(Y_2)\big) \mid R^X(X_1)=x_1, R^X(X_2)=x_2\big]$ is uniformly β-Hölder continuous w.r.t. x_1,x_2. Applying this to (5.3) shows

$$\limsup_{n\to\infty} \text{r.h.s. of (5.3)}$$
$$\lesssim \limsup_{n\to\infty}\frac{\max_{j=1}^n d_j}{n\min_{j=1}^n d_j}\sum_{i=1}^n\frac{1}{d_i}\mathbb{E}\Big[\sum_{k:(k,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \big\|R^X(X_i)-R^X(X_k)\big\|^\beta\Big]$$
$$\leq \limsup_{n\to\infty}\frac{\max_{j=1}^n d_j}{n\min_{j=1}^n d_j}\mathbb{E}\Big[\sum_{i=1}^n\frac{1}{d_i}\sum_{k:(k,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \big\|\widehat{R}_n^X(X_i)-\widehat{R}_n^X(X_k)\big\|^\beta\Big]. \tag{5.4}$$

where the last inequality follows once again from Proposition 5.1. Notice that $\sum_{i=1}\sum_{k:(k,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})}\big\|\widehat{R}_n^X(X_i)-\widehat{R}_n^X(X_k)\big\|^\beta = \sum_{e\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})}|e|^\beta$ where $|e|$ denotes the length

of the edge e. The right hand side of the above display converge to 0 by (3.5). This shows $|(\mathbf{III}) - (\widetilde{\mathbf{III}})|$ converges to 0 as n tends to ∞. On other hand, we define

$$
(\widetilde{\mathbf{I}}) := \frac{1}{n^2}\mathbb{E}\Bigg[\sum_{i=1}^{n}\sum_{j:(j,i)\in\mathcal{E}(\mathcal{G}_n^{\mathrm{RANK}})}\bigg(\frac{1}{d_i^2}\Big[\big(K(R^Y(Y_i), R^Y(Y_i'))\big)^2\Big]
$$
$$
+ \frac{1}{d_i d_j}\mathbb{E}\Big[K(R^Y(Y_i), R^Y(Y_i'))\, K(R^Y(Y_j), R^Y(Y_j'))\Big]\bigg)\Bigg],
$$

and

$$
(\widetilde{\mathbf{II}}) := \frac{1}{n^2}\mathbb{E}\Bigg[\sum_{i=1}^{n}\sum_{(j,i),(k,i)\in\mathcal{E}(\mathcal{G}_n^{\mathrm{RANK}})}\bigg[\frac{1}{d_j d_k}K(R^Y(Y_k), R^Y(Y_k'))\, K(R^Y(Y_j), R^Y(Y_j'))
$$
$$
+ \frac{1}{d_i^2}K^2(R^Y(Y_i), R^Y(Y_i')) + \frac{2}{d_i d_j}K(R^Y(Y_i), R^Y(Y_i'))\, K(R^Y(Y_j), R^Y(Y_j'))\bigg]\Bigg].
$$

Note that

$$
(\widetilde{\mathbf{I}}) + (\widetilde{\mathbf{II}}) + (\widetilde{\mathbf{III}}) = \frac{1}{n^2}\mathbb{E}\Bigg[\Big(\sum_{i=1}^{n}K(R^Y(Y_i), R^Y(Y_i'))\Big)^2\Bigg]
$$
$$
\to \Big(\mathbb{E}\big[K(R^Y(Y_1), R^Y(Y_1'))\big]\Big)^2 \tag{5.5}
$$

where the last convergence follows from the strong law of large numbers and the dominated convergence theorem. Owing to the fact that (3.4) holds and $K(\cdot,\cdot)$ is continuous a.e., it is easy to check that $(\mathbf{I}), (\mathbf{II}), (\widetilde{\mathbf{I}}), (\widetilde{\mathbf{II}})$ are converging to 0 as $n \to \infty$. Combining this with (5.2) and (5.5) completes the proof of (a).

Now we move on to (b). By the towering property of the conditional expectation, we have

$$
\mathbb{E}\mathcal{Z}_n = \mathbb{E}\Big[\frac{1}{n}\sum_i d_i^{-1}\sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)} K(R^Y(Y_i), R^Y(Y_j))\Big]
$$
$$
= \mathbb{E}\Big[\frac{1}{n}\sum_i d_i^{-1}\sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)} \mathbb{E}\big[K(R^Y(Y_i), R^Y(Y_j)) \mid R^X(X_i), R^X(X_j)\big]\Big]. \tag{5.6}
$$

By the β-Hölder continuity of $\mathbb{E}\big[K(R^Y(Y_i), R^Y(Y_j)) \mid R^X(X_i) = x, R^X(X_j) = y\big]$ as a function of x and y (see (3.3)), there exists $C > 0$ such that

$$
\bigg|\mathbb{E}\big[K(R^Y(Y_i), R^Y(Y_j)) \mid R^X(X_i), R^X(X_j)\big]
$$
$$
- \mathbb{E}\big[K(R^Y(Y_i), R^Y(Y_j')) \mid R^X(X_i)\big]\bigg| \lesssim \|R^X(X_i) - R^X(X_j)\|^{\beta}
$$

where Y_i, Y_i' are two independent samples from the conditional distribution of Y given X_i. Applying the above inequality, we get

$$
\bigg|\text{r.h.s. of (5.6)} - \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} K(R^Y(Y_i), R^Y(Y_i'))\Big]\bigg|
$$

$$\lesssim \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\sum_{j:(i,j)\in\mathcal{E}(\mathcal{G}_n)}\frac{1}{d_i}\|R^X(X_i)-R^X(X_j)\|^{\beta}\Big].$$

The proof of (b) can now be completed using the same steps as those in the proof of (a) starting from (5.4). □

5.2. Proof of Theorem 3.2

Proof of (P1): This follows from [22, Theorem 2.1].

Proof of (P2): By Proposition 2.1, there exists $Q^Y(\cdot)$ such that $Q^Y(R^Y(y)) = y$ a.e. μ_Y. By [22, Theorem 2.1], it follows that $\eta_K^{\mathrm{RANK}} = 0$ if and only if $R^Y(Y)$ and X are independent. As $Q^Y(R^Y(Y)) = Y$, we further have that $R^Y(Y)$ and X are independent if and only if Y and X are independent. The conclusion then follows.

Proof of (P3): By [22, Theorem 2.1], it follows that $\eta_K^{\mathrm{RANK}} = 1$ if and only if there exists a measurable function $h(\cot)$ such that $R^Y(Y) = h(X)$ a.e. μ. Now $R^Y(Y) = h(X) \Leftrightarrow Y = Q^Y(h(X))$. The conclusion then follows by choosing $g \equiv Q^Y \circ h$. □

5.3. Proof of Proposition 3.1

The crucial observation in this proof is that $K(y_1, y_2) = |y_1| + |y_2| - |y_1 - y_2| = \min\{y_1, y_2\}$ for $y_1, y_2 \in [0, \infty)$. Now, when $d_2 = 1$, $R^Y(\cdot)$ is simply the cumulative distribution function of Y; we will call it F_Y to stick with conventional notation. Next note that, $\eta_K^{\mathrm{RANK}}(\mu)$ can be simplified as:

$$\eta_K^{\mathrm{RANK}}(\mu) = \frac{\mathbb{E}\min\{F_Y(Y'), F_Y(\widetilde{Y'})\} - \mathbb{E}\min\{F_Y(Y_1), F_Y(Y_2)\}}{1/2 - \mathbb{E}\min\{F_Y(Y_1), F_Y(Y_2)\}}. \tag{5.7}$$

As $\min\{a, b\} = \int_0^\infty \mathbf{1}(t \le a)\mathbf{1}(t \le b)\, dt$ for $a, b \in \mathbb{R}$, an application of the dominated convergence theorem yields:

$$\mathbb{E}\min\{F_Y(Y'), F_Y(\widetilde{Y'})\} = \int \mathbb{E}[(\mathbb{P}(F_Y(Y) \ge t|X))^2]\, dt = \int \mathbb{E}[(\mathbb{P}(Y \ge t|X))^2]\, d\mu_Y(t)$$

and similarly,

$$\mathbb{E}\min\{F_Y(Y_1), F_Y(Y_2)\} = \int (\mathbb{P}(Y \ge t))^2\, d\mu_Y(t).$$

Plugging the above expressions into the expression of $\eta_K^{\mathrm{RANK}}(\mu)$ in (5.7) gives:

$$\eta_K^{\mathrm{RANK}}(\mu) = \frac{\int \Big(\mathbb{E}[(\mathbb{P}(Y \ge t|X))^2] - (\mathbb{P}(Y \ge t))^2\Big)\, d\mu_Y(t)}{\int \Big(\mathbb{P}(Y \ge t) - (\mathbb{P}(Y \ge t))^2\Big)\, d\mu_Y(t)} = \xi(\mu).$$

This completes the proof. □

5.4. Proof of Theorem 3.3

Note the definitions of $\widetilde{a}, \widetilde{b}, \widetilde{c}, \widetilde{g}_1, \widetilde{g}_2, \widetilde{g}_3$ from the statement of Theorem 3.3. For this proof, we will also borrow some elements used in the proof of [22, Theorem 4.1]. Define $C_n := (n(n - $

$1))^{-1} \sum_{i \neq j} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j))$ and recall that C_n is a deterministic quantity as was explained in the comment after (5.1). Also $\mathbb{E}K(\widehat{R}_n^Y(Y_1), \widehat{R}_n^Y(Y_2)) = C_n$. Let $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, i.e., the σ-field generated by $(X_1, \ldots, X_n)$. Consequently, note that:

$$\mathbb{E}[N_n^{\text{RANK}}|\mathcal{F}_n] = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j:(j,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} \mathbb{E}[K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j))] - C_n\right)$$

$$= \sqrt{n}C_n\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j:(j,i)\in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})} 1 - 1\right) = 0. \tag{5.8}$$

By the same calculation as in [22, Equation (C.12)], we get:

$$\text{Var}(N_n^{\text{RANK}}|\mathcal{F}_n) = \left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1}\right)(\widehat{a} - 2\widehat{b} + \widehat{c}) + (\widetilde{g}_2 - 1)(\widehat{b} - \widehat{c}), \tag{5.9}$$

where

$$\widehat{a} := \frac{1}{n(n-1)} \sum_{(i,j) \text{ distinct}} K^2(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j))$$

$$\widehat{b} := \frac{1}{n(n-1)(n-2)} \sum_{(i,j,l) \text{ distinct}} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j))K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_l))$$

$$\widehat{c} := \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,l,m) \text{ distinct}} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j))K(\widehat{R}_n^Y(Y_l), \widehat{R}_n^Y(Y_m)).$$

Now $\widehat{a}, \widehat{b}, \widehat{c}$ are clearly $\mathcal{O}(1)$ and $\widetilde{g}_1, \widetilde{g}_2, \widetilde{g}_3$ are $\mathcal{O}(1)$ by using [22, Equation (C.13)]. Next observe that the right hand side of (5.9) is a deterministic quantity. Therefore by combining (5.8) and (5.9), we get $\text{Var}(N_n^{\text{RANK}}) = \mathcal{O}(1)$.

In order to establish the CLT, define

$$N_n^{\text{POP}} := \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{\substack{j:(j,i) \\ \in\mathcal{E}(\mathcal{G}_n^{\text{RANK}})}} K(R^Y(Y_i), R^Y(Y_j)) - \frac{\sum_{i\neq j} K(R^Y(Y_i), R^Y(Y_j))}{n(n-1)}\right). \tag{5.10}$$

It then suffices to show the following:

$$\sup_{\widetilde{\mathcal{G}}\in\mathcal{J}_\theta}\left|\frac{\text{Var}(N_n^{\text{POP}})}{\text{Var}(N_n^{\text{RANK}})} - 1\right| \overset{n\to\infty}{\longrightarrow} 0, \tag{5.11}$$

$$\sup_{\widetilde{\mathcal{G}}\in\mathcal{J}_\theta} \mathbb{E}\left[(N_n^{\text{POP}} - N_n^{\text{RANK}})^2\right] \overset{n\to\infty}{\longrightarrow} 0, \tag{5.12}$$

$$\sup_{\widetilde{\mathcal{G}}\in\mathcal{J}_\theta} \sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{N_n^{\text{POP}}}{\sqrt{\text{Var}(N_n^{\text{POP}})}} \leq x\right) - \Phi(x)\right| \overset{n\to\infty}{\longrightarrow} 0. \tag{5.13}$$

Proof of (5.11). As in [22, Equation (C.12)], we have:

$$\mathrm{Var}(N_n^{\mathrm{POP}}) = \mathbb{E}(N_n^{\mathrm{POP}})^2 = \left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1}\right)(\widetilde{a} - 2\widetilde{b} + \widetilde{c}) + (\widetilde{g}_2 - 1)(\widetilde{b} - \widetilde{c}). \qquad (5.14)$$

Therefore,

$$\sup_{\widetilde{\mathcal{G}} \in \mathcal{J}_\theta} \left| \frac{\mathrm{Var}(N_n^{\mathrm{POP}})}{\mathrm{Var}(N_n^{\mathrm{RANK}})} - 1 \right| \lesssim \sup_{\widetilde{\mathcal{G}} \in \mathcal{J}_\theta} \left[\left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1}\right)(\widetilde{a} - \widehat{a} - 2\widetilde{b} + 2\widehat{b} + \widetilde{c} - \widehat{c}) \right.$$
$$\left. + (\widetilde{g}_2 - 1)(\widetilde{b} - \widehat{b} - \widetilde{c} + \widehat{c}) \right] \overset{n \to \infty}{\longrightarrow} 0.$$

The conclusion above follows by noting that $\widehat{a} \to \widetilde{a}$, $\widehat{b} \to \widetilde{b}$, and $\widehat{c} \to \widetilde{c}$, as $n \to \infty$, by assumption (S3).

Proof of (5.12). Let us first introduce the notion of the *resampling distribution*. Note that when $\widetilde{\mu} = \widetilde{\mu}_X \otimes \widetilde{\mu}_Y$, the joint distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$ is the same as the joint distribution of $(X_1, Y_{\sigma(1)}), \ldots, (X_n, Y_{\sigma(n)})$ where σ is a random permutation of the set $\{1, 2, \ldots, n\}$ which is drawn independent of the X_i's and Y_i's.

The expressions for $\mathbb{E}(N_n^{\mathrm{RANK}})^2$ and $\mathbb{E}(N_n^{\mathrm{POP}})^2$ has already been presented in (5.9) and (5.14) respectively. Therefore, in the sequel, we will only focus on the term $\mathbb{E}[N_n^{\mathrm{RANK}} N_n^{\mathrm{POP}}]$ where we will use the *resampling distribution* as discussed above. Towards this direction, let us define:

$$\underline{a} := \frac{1}{n(n-1)} \sum_{(i,j) \text{ distinct}} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) K(R^Y(Y_i), R^Y(Y_j))$$

$$\underline{b} := \frac{1}{n(n-1)(n-2)} \sum_{(i,j,l) \text{ distinct}} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) K(R^Y(Y_i), R^Y(Y_l))$$

$$\underline{c} := \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,l,m) \text{ distinct}} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) K(R^Y(Y_l), R^Y(Y_m)).$$

Also to simplify notation, we will use the symbol $j \sim i$ for $(j, i) \in \mathcal{E}(\mathcal{G}_n^{\mathrm{RANK}})$ and let $\widehat{Z}_i := \widehat{R}_n^Y(Y_i)$ and $Z_i := R^Y(Y_i)$. With the above notation, observe that $\mathbb{E}[N_n^{\mathrm{RANK}} N_n^{\mathrm{POP}}]$ can be simplified as follows:

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j \sim i} K(\widehat{R}_n^Y(Y_{\sigma(i)}), \widehat{R}_n^Y(Y_{\sigma(j)})) \right) \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j \sim i} K(R^Y(Y_{\sigma(i)}), R^Y(Y_{\sigma(j)})) \right) \right.$$

$$- \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j \sim i} K(\widehat{R}_n^Y(Y_{\sigma(i)}), \widehat{R}_n^Y(Y_{\sigma(j)})) \right) \left(\frac{1}{n(n-1)} \sum_{i \neq j} K(R^Y(Y_i), R^Y(Y_j)) \right)$$

$$- \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{d}_i} \sum_{j \sim i} K(R^Y(Y_{\sigma(i)}), R^Y(Y_{\sigma(j)})) \right) \left(\frac{1}{n(n-1)} \sum_{i \neq j} K(\widehat{R}_n^Y(Y_i), \widehat{R}_n^Y(Y_j)) \right)$$

$$\left. + \left(\frac{1}{n(n-1)} \sum_{i \neq j} K(\widehat{R}_n^Y(Y_{\sigma(i)}), \widehat{R}_n^Y(Y_{\sigma(j)})) \right) \left(\frac{1}{n(n-1)} \sum_{i \neq j} K(R^Y(Y_{\sigma(i)}), R^Y(Y_{\sigma(j)})) \right) \right]$$

$$
\overset{(*)}{=} (\widetilde{g}_1 + \widetilde{g}_3)\mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})K(Z_{\sigma(1)}, Z_{\sigma(2)})\right] + (3 + \widetilde{g}_2 - 2\widetilde{g}_1 - 2\widetilde{g}_3)\mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})\right.
$$

$$
\left. K(Z_{\sigma(1)}, Z_{\sigma(3)})\right] + (n - 3 + \widetilde{g}_1 + \widetilde{g}_2 + \widetilde{g}_3)\mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})K(Z_{\sigma(3)}, Z_{\sigma(4)})\right]
$$

$$
- \frac{2}{n-1} \cdot \mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})K(Z_{\sigma(1)}, Z_{\sigma(2)})\right] - \frac{4(n-2)}{n-1} \cdot \mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})K(Z_{\sigma(1)}, Z_{\sigma(3)})\right]
$$

$$
- \frac{(n-2)(n-3)}{n-1} \cdot \mathbb{E}\left[K(\widehat{Z}_{\sigma(1)}, \widehat{Z}_{\sigma(2)})K(Z_{\sigma(3)}, Z_{\sigma(4)})\right]
$$

$$
= \left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1}\right)(\underline{a} - 2\underline{b} + \underline{c}) + (\widetilde{g}_2 - 1)(\underline{b} - \underline{c}).
$$

Here $(*)$ follows from [22, Equations (C.9), (C.10), and (C.11)]. By Assumption (S3), $\widehat{a}, \underline{a}$ converge to $\widetilde{a}$ in L^1; same holds for $\widehat{b}, \underline{b}, \widetilde{b}$ and $\widehat{c}, \underline{c}, \widetilde{c}$. Also $\widehat{a}, \underline{a}, \widetilde{a}, \widehat{b}, \underline{b}, \widetilde{b}, \widehat{c}, \underline{c}, \widetilde{c}$ do not depend on the graph functional $\mathcal{G}$. Therefore, the following holds:

$$
\sup_{\widetilde{\mathcal{G}} \in \mathcal{J}_\theta} \mathbb{E}\left[(N_n^{\mathrm{RANK}} - N_n^{\mathrm{POP}})\right]^2 = \sup_{\widetilde{\mathcal{G}} \in \mathcal{J}_\theta} \left[\left(\widetilde{g}_1 + \widetilde{g}_3 - \frac{2}{n-1}\right)\{\widehat{a} - 2\underline{a} + \widetilde{a} + 4\underline{b} - 2\widetilde{b}\right.
$$

$$
\left. - 2\widehat{b} + \widehat{c} - 2\underline{c} + \widetilde{c}\} + (\widetilde{g}_2 - 1)(\widehat{b} - 2\underline{b} + \widetilde{b} - \widehat{c} + 2\underline{c} - \widetilde{c})\right] \overset{n\to\infty}{\longrightarrow} 0.
$$

This completes the proof of (5.12).

Proof of (5.13). The conclusion follows directly from [22, Theorem 4.1]. $\qquad\square$

References

[1] Akhiezer, N. I. and I. M. Glazman (1993). *Theory of linear operators in Hilbert space.* Dover Publications, Inc., New York. Translated from the Russian and with a preface by Merlynd Nestell, Reprint of the 1961 and 1963 translations, Two volumes bound as one.

[2] Ansari, J. and S. Fuchs (2022). A simple extension of Azadkia & Chatterjee's rank correlation to a vector of endogenous variables. *arXiv preprint arXiv:2212.01621*.

[3] Auddy, A., N. Deb, and S. Nandy (2024). Exact detection thresholds and minimax optimality of Chatterjee's correlation coefficient. *Bernoulli 30*(2), 1640–1668.

[4] Azadkia, M. and S. Chatterjee (2021). A simple measure of conditional dependence. *Ann. Statist. 49*(6), 3070–3102.

[5] Berlinet, A. and C. Thomas-Agnan (2004). *Reproducing kernel Hilbert spaces in probability and statistics.* Kluwer Academic Publishers, Boston, MA. With a preface by Persi Diaconis.

[6] Berrett, T. B. and R. J. Samworth (2019). Nonparametric independence testing via mutual information. *Biometrika 106*(3), 547–566.

[7] Berrett, T. B. and R. J. Samworth (2023). Efficient functional estimation and the super-oracle phenomenon. *Ann. Statist. 51*(2), 668–690.

[8] Berrett, T. B., R. J. Samworth, and M. Yuan (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist. 47*(1), 288–318.

[9] Bertsekas, D. P. (1988). The auction algorithm: a distributed relaxation method for the assignment problem. *Ann. Oper. Res. 14*(1-4), 105–123.

[10] Bhattacharya, B. B. (2020). Asymptotic distribution and detection thresholds for two-sample tests based on geometric graphs. *Ann. Statist. 48*(5), 2879–2903.

[11] Biswas, M., S. Sarkar, and A. K. Ghosh (2016). On some exact distribution-free tests of independence between two random vectors of arbitrary dimensions. *J. Statist. Plann. Inference 175*, 78–86.

[12] Blum, J. R., J. Kiefer, and M. Rosenblatt (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist. 32*, 485–498.

[13] Boonmee, T. and S. Tasena (2016). Measure of complete dependence of random vectors. *J. Math. Anal. Appl. 443*(1), 585–595.

[14] Cao, S. and P. J. Bickel (2020). Correlations with tailored extremal properties. *arXiv preprint arXiv:2008.10177*.

[15] Chattamvelli, R. (2024). Measures of association. In *Correlation in Engineering and the Applied Sciences: Applications in R*, pp. 1–54. Springer.

[16] Chatterjee, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc. 116*(536), 2009–2022.

[17] Chatterjee, S. (2024). *A Survey of Some Recent Developments in Measures of Association*, pp. 109–128. Singapore: Springer Nature Singapore.

[18] Chernoff, H. and I. R. Savage (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist. 29*, 972–994.

[19] Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist. 45*(1), 223–256.

[20] Cordero-Erausquin, D. and A. Figalli (2019). Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst. 39*(12), 7101–7112.

[21] Deb, N., B. B. Bhattacharya, and B. Sen (2021). Pitman efficiency lower bounds for multivariate distribution-free tests based on optimal transport. *arXiv preprint arXiv:2104.01986*.

[22] Deb, N., P. Ghosal, and B. Sen (2020). Measuring association on topological spaces using kernels and geometric graphs. *arXiv preprint arXiv:2010.01768*.

[23] Deb, N. and B. Sen (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc. 118*(541), 192–207.

[24] Dette, H. and M. Kroll (2024). A simple bootstrap for Chatterjee's rank correlation. *Biometrika*.

[25] Dette, H., K. F. Siburg, and P. A. Stoimenov (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat. 40*(1), 21–41.

[26] Diestel, J. and B. Faires (1974). On vector measures. *Trans. Amer. Math. Soc. 198*, 253–271.

[27] Dinculeanu, N. (2011). Vector integration and stochastic integration in Banach spaces. In *The Oxford handbook of functional data analysis*, pp. 327–354. Oxford Univ. Press, Oxford.

[28] Faden, A. M. (1985). The existence of regular conditional probabilities: necessary and sufficient conditions. *Ann. Probab. 13*(1), 288–298.

[29] Friedman, J. H. and L. C. Rafsky (1983). Graph-theoretic measures of multivariate association and prediction. *Ann. Statist. 11*(2), 377–391.

[30] Fuchs, S. (2024). Quantifying directed dependence via dimension reduction. *J. Multivariate Anal. 201*, Paper No. 105266.

[31] Gieser, P. W. and R. H. Randles (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc. 92*(438), 561–567.

[32] Gretton, A., K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola (2012). A kernel two-sample test. *Journal of Machine Learning Research 13*(Mar), 723–773.

[33] Gretton, A., K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola (2008).

A kernel statistical test of independence. In *Advances in neural information processing systems*, pp. 585–592.

[34] Gretton, A. and L. Györfi (2008). Nonparametric independence tests: space partitioning and kernel approaches. In *Algorithmic learning theory*, Volume 5254 of *Lecture Notes in Comput. Sci.*, pp. 183–198. Springer, Berlin.

[35] Hallin, M., E. del Barrio, J. Cuesta-Albertos, and C. Matrán (2021). Distribution and quantile functions, ranks and signs in dimension d: a measure transportation approach. *Ann. Statist. 49*(2), 1139–1165.

[36] Hallin, M., D. Hlubinka, and Š. Hudecová (2023). Efficient fully distribution-free center-outward rank tests for multiple-output regression and manova. *J. Amer. Statist. Assoc. 118*(543), 1923–1939.

[37] Heller, R., Y. Heller, and M. Gorfine (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika 100*(2), 503–510.

[38] Hodges, Jr., J. L. and E. L. Lehmann (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist. 27*, 324–335.

[39] Hofer, R. (2009). On the distribution properties of Niederreiter-Halton sequences. *J. Number Theory 129*(2), 451–463.

[40] Hofer, R. and G. Larcher (2010). On existence and discrepancy of certain digital Niederreiter-Halton sequences. *Acta Arith. 141*(4), 369–394.

[41] Huber, P. J. and E. M. Ronchetti (2009). *Robust statistics* (Second ed.). Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

[42] Josse, J. and S. Holmes (2016). Measuring multivariate association and beyond. *Stat. Surv. 10*, 132–167.

[43] Kendall, M. and J. D. Gibbons (1990). *Rank correlation methods* (Fifth ed.). A Charles Griffin Title. Edward Arnold, London.

[44] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika 30*(1/2), 81–93.

[45] Lin, Z. and F. Han (2023). On boosting the power of Chatterjee's rank correlation. *Biometrika 110*(2), 283–299.

[46] Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab. 41*(5), 3284–3305.

[47] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J. 80*(2), 309–323.

[48] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mémoires Acad. Royale Sci. 1781*, 666–704.

[49] Muandet, K., K. Fukumizu, B. Sriperumbudur, B. Schölkopf, et al. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning 10*(1-2), 1–141.

[50] Munkres, J. (1957). Algorithms for the assignment and transportation problems. *J. Soc. Indust. Appl. Math. 5*, 32–38.

[51] Nies, T. G., T. Staudt, and A. Munk (2021). Transport dependency: Optimal transport based dependency measures. *arXiv preprint arXiv:2105.02073*.

[52] Oja, H. (2010). *Multivariate nonparametric methods with R*, Volume 199 of *Lecture Notes in Statistics*. Springer, New York. An approach based on spatial signs and ranks.

[53] Pearson, K. (1920). Notes on the history of correlation. *Biometrika 13*(1), 25–45.

[54] Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar. 10*, 441–451.

[55] Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti (2011). Detecting novel associations in large data sets. *Science 334*(6062), 1518–1524.

[56] Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist. 3*, 1–14.

[57] Sen, A. and B. Sen (2014). Testing independence and goodness-of-fit in linear models. *Biometrika 101*(4), 927–942.

[58] Shi, H., M. Drton, M. Hallin, and F. Han (2025). Distribution-free tests of multivariate independence based on center-outward quadrant, Spearman, Kendall, and van der Waerden statistics. *Bernoulli 31*(1), 106 – 129.

[59] Shi, H., M. Drton, and F. Han (2022a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc. 117*(537), 395–410.

[60] Shi, H., M. Drton, and F. Han (2022b). On the power of Chatterjee's rank correlation. *Biometrika 109*(2), 317–333.

[61] Shi, H., M. Hallin, M. Drton, and F. Han (2022). On universally consistent and fully distribution-free rank tests of vector independence. *Ann. Statist. 50*(4), 1933–1959.

[62] Siburg, K. F. and P. A. Stoimenov (2010). A measure of mutual complete dependence. *Metrika 71*(2), 239–251.

[63] Smola, A., A. Gretton, L. Song, and B. Schölkopf (2007). A Hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pp. 13–31. Springer.

[64] Spearman, C. (1904). The proof and measurement of association between two things. *American journal of Psychology 15*(1), 72–101.

[65] Steinwart, I. and A. Christmann (2008). *Support vector machines*. Information Science and Statistics. Springer, New York.

[66] Strothmann, C., H. Dette, and K. F. Siburg (2024). Rearranged dependence measures. *Bernoulli 30*(2), 1055–1078.

[67] Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist. 35*(6), 2769–2794.

[68] Tasena, S. and S. Dhompongsa (2016). Measures of the functional dependence of random vectors. *Internat. J. Approx. Reason. 68*, 15–26.

[69] Taskinen, S., H. Oja, and R. H. Randles (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc. 100*(471), 916–925.

[70] van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

[71] Villani, C. (2003). *Topics in optimal transportation*, Volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

[72] Zhang, Q. (2023). On the asymptotic null distribution of the symmetrized Chatterjee's correlation coefficient. *Statist. Probab. Lett. 194*, Paper No. 109759, 7.