Electronic Journal of Statistics
Vol. 10 (2016) 2287–2311
ISSN: 1935-7524
DOI: 10.1214/16-EJS1166

Large-sample theory for the Bergsma-Dassios sign covariance

Preetam Nandy[*,†]

Seminar for Statistics, ETH Zürich, Switzerland
e-mail: nandy@stat.math.ethz.ch

Luca Weihs[*] **and Mathias Drton**

Department of Statistics, University of Washington, Seattle, WA, USA
e-mail: lucaw@uw.edu; md5@uw.edu

Abstract: The Bergsma-Dassios sign covariance is a recently proposed extension of Kendall's tau. In contrast to tau or also Spearman's rho, the new sign covariance τ^* vanishes if and only if the two considered random variables are independent. Specifically, this result has been shown for continuous as well as discrete variables. We develop large-sample distribution theory for the empirical version of τ^*. In particular, we use theory for degenerate U-statistics to derive asymptotic null distributions under independence and demonstrate in simulations that the limiting distributions give useful approximations.

MSC 2010 subject classifications: Primary 62G10, 62G20; secondary 62G30.
Keywords and phrases: Test of independence, asymptotics, U-statistics, nonparametric correlation, degeneracy, Hoeffding's D.

Received February 2016.

1. Introduction

Many popular measures of pairwise dependence, for example Kendall's tau [7] and Spearman's rho [11], have the undesirable property that they may be zero even when the two considered random variables X and Y are dependent. Addressing this weakness, Bergsma and Dassios [1] introduced a new rank-based correlation measure τ^*, which, under mild conditions on the joint distribution of (X, Y), is zero if and only if X and Y are independent. Where Kendall's tau is defined in terms of concordance and discordance of two independent copies of (X, Y), the new τ^* is based on similar notions of concordance and discordance for four independent copies of (X, Y). While a naïve computation of t^*, the empirical version of τ^*, thus requires $O(n^4)$ time for a sample of size n, it was recently shown that this computational burden can be reduced to $O(n^2)$ [5, 15]. As t^* is now computable for larger sample sizes, understanding its asymptotic

*Equal contribution.
†Supported in part by Swiss NSF Grant 200021_149760.

behavior becomes a problem of practical interest and has the potential to yield simple tests of independence that avoid Monte Carlo approximation of p-values.

We introduce the statistic t^* in Section 2, where we also review background on U-statistics. In Section 3, we clarify that t^* is a degenerate U-statistic under the null hypothesis that the sample is generated under independence. We also prove that in certain settings degeneracy occurs only under independence. In Section 4, we use the asymptotic theory of degenerate U-statistics to derive an explicit representation of the asymptotic distribution of t^* when the sample is generated under independence and with marginals that are continuous or discrete. The asymptotic distribution takes the form of a Gaussian chaos; specifically, we find a (in some cases infinite) sum of scaled and centered chi-square distributions. Our asymptotic results are deeply connected to the notion of a Hoeffding decomposition, a technique which breaks down a random variable into a sum of projections onto carefully constructed orthogonal spaces [13, Section 11.4]. In this setting, the degeneracy of t^* under independence is equivalent to its projection upon one of these spaces equalling 0. We note that the asymptotics of U-statistics also have interesting connections to stochastic integration, indeed the theorems of Section 2.2 can be understood using representations of Wiener integrals [4]. Simulations in Section 5 demonstrate how the large-sample theory can be leveraged to perform tests of independence and compute power. Indeed, asymptotic distributions are found to give accurate approximations for sample sizes as small as $n = 80$. We end with a discussion in Section 6.

2. Preliminaries

2.1. The t^* statistic

Let $(x_1, y_1), \dots, (x_n, y_n)$ be a sample of points in $\mathbb{R}^2$. The empirical version of the Bergsma-Dassios sign covariance is the statistic

$$t^* := \frac{(n-4)!}{n!} \sum_{\substack{1\le i,j,k,l\le n \\ i,j,k,l \text{ distinct}}} a(x_i, x_j, x_k, x_l) a(y_i, y_j, y_k, y_l), \tag{1}$$

where

$$\begin{aligned} &a(z_1, z_2, z_3, z_4) \\ &\quad = I(z_1, z_3 < z_2, z_4) + I(z_1, z_3 > z_2, z_4) - I(z_1, z_2 < z_3, z_4) - I(z_1, z_2 > z_3, z_4). \end{aligned} \tag{2}$$

Here we use $I(\cdot)$ to denote the indicator function and $a, b < c, d$ is shorthand for $\max(a, b) < \min(c, d)$. As in [15], we defined t^* in the form of a U-statistic, whereas in [1] it is introduced as a V-statistic. Indeed, t^* from (1) is an unbiased estimator of the sign covariance

$$\tau^* := E\left[a(X_1, X_2, X_3, X_4) a(Y_1, Y_2, Y_3, Y_4)\right]$$

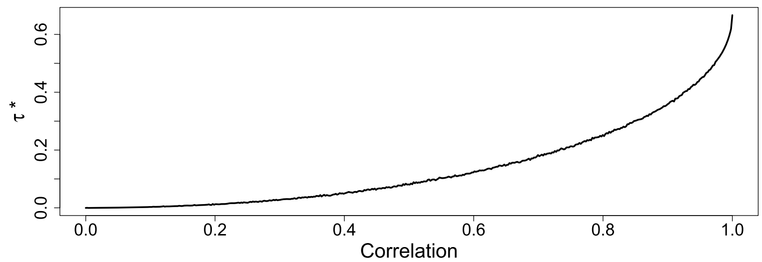


FIG 1. *Sign covariance τ^* of bivariate normal distributions.*

of [1]. Here, $(X_1, Y_1), \dots, (X_4, Y_4)$ are random vectors drawn independently from a given bivariate distribution on $\mathbb{R}^2$.

Example 1. *Figure 1 shows the values of τ^* for bivariate normal distributions, which we computed by Monte Carlo simulation. The sign covariance τ^* is an even function of the normal correlation ρ, and we thus only show values for $\rho \in [0, 1]$. For each considered correlation ρ we averaged 200 values of t^*, each computed from a sample of size $n = 300$.*

As we explain in the remainder of this subsection, the statistic t^* is based on counting concordant and disconcordant quadruples.

Definition 2.1. *Let $(x_1, y_1), \dots, (x_4, y_4)$ be four points relabelled so that $x_1 \le x_2 \le x_3 \le x_4$. We say that the points are*

inseparable *if* $x_2 = x_3$ *or there exists a permutation* π *of* $\{1, 2, 3, 4\}$ *so that* $y_{\pi(1)} \le y_{\pi(2)} = y_{\pi(3)} \le y_{\pi(4)}$,

and if they are not inseparable, then we call them

concordant *if* $\max(y_1, y_2) < \min(y_3, y_4)$ *or* $\max(y_3, y_4) < \min(y_1, y_2)$,
discordant *if* $\max(y_1, y_2) > \min(y_3, y_4)$ *and* $\max(y_3, y_4) > \min(y_1, y_2)$.

The above definitions are mutually exclusive and exhaustive in that any set of four points in $\mathbb{R}^2$ will be exactly one of inseparable, concordant, or discordant. Moreover, if the points are drawn from a bivariate distribution with continuous marginals then they will be almost surely concordant or discordant. See Figure 3 of [1] for a visual depiction of concordance and discordance.

Let S_4 be the set of permutations on 4 elements, and for $\pi \in S_4$ and $(z_1, z_2, z_3, z_4) \in \mathbb{R}^4$, let $z_{\pi(1,2,3,4)} := (z_{\pi(1)}, z_{\pi(2)}, z_{\pi(3)}, z_{\pi(4)})$. Introducing the symmetric function

$$h((x_1, y_1), \dots, (x_4, y_4)) := \frac{1}{4!} \sum_{\pi \in S_4} a(x_{\pi(1,2,3,4)}) a(y_{\pi(1,2,3,4)}), \tag{3}$$

we may rewrite t^* as a sum of permutation invariant terms, namely,

$$t^* = \frac{1}{\binom{n}{4}} \sum_{(i,j,k,l) \in C(n,4)} h\left((x_i, y_i), (x_j, y_j), (x_k, y_k), (x_l, y_l)\right), \tag{4}$$

where $C(n,4) = \{(i,j,k,l) : 1 \le i < j < k < l \le n\}$. Lemma 1 in [15] gives the following result.

Lemma 2.1. *Let* $A = \{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)\} \subset \mathbb{R}^2$. *Then*

$$h((x_1,y_1),..,(x_4,y_4)) = \begin{cases} 2/3 & \textit{if the points in } A \textit{ are concordant,} \\ -1/3 & \textit{if the points in } A \textit{ are discordant,} \\ 0 & \textit{if the points in } A \textit{ are inseparable.} \end{cases}$$

Equation (4) expresses t^* in the familiar form of a U-statistic with symmetric kernel h, and we proceed to review some of the tools available for the study of U-statistics.

2.2. Theory of U-statistics

Let $Z_1, Z_2, \dots$ be i.i.d. random variables taking their values in $\mathbb{R}^d$ with $d \ge 1$. Let $k : \left(\mathbb{R}^d\right)^m \to \mathbb{R}$ be a *kernel function* invariant to permutation of its m arguments. For $n \ge m$, the *U-statistic with kernel* k is the statistic

$$U_n := \frac{1}{\binom{n}{m}} \sum_{(i_1,\dots,i_m)\in C(n,m)} k(Z_{i_1},\dots,Z_{i_m}), \tag{5}$$

where $C(n,m) = \{(i_1,\dots,i_m) \in \{1,\dots,n\}^m : i_1 < i_2 < \dots < i_m\}$. Note that $E[U_n] = E[k(Z_1,\dots,Z_m)]$ so that U_n is an unbiased estimator of $\theta := E[k(Z_1,\dots,Z_m)]$.

Of central importance in determining the asymptotics of U-statistics are the functions

$$k_i(z_1,\dots,z_i) = E[k(z_1,\dots,z_i,Z_{i+1},\dots,Z_m)], \quad i = 1,\dots,m, \tag{6}$$

and their variances

$$\sigma_i^2 = \mathrm{Var}[k_i(Z_1,\dots,Z_i)], \quad i = 1,\dots,m. \tag{7}$$

It is well known that $\sigma_1^2 \le \sigma_2^2 \le \dots \le \sigma_m^2$. In particular, if σ_m^2 is finite then so are all other σ_i^2. We now recall two theorems on the large-sample distribution of the U-statistic U_n [10, Chapter 5].

Theorem 2.2. *If the kernel* k *of the statistic* U_n *from (5) has variance* $\sigma_m^2 < \infty$, *then*

$$\sqrt{n}(U_n - \theta) \xrightarrow{d} N(0, m^2\sigma_1^2).$$

If $\sigma_1^2 = 0$, then the Gaussian limit is degenerate, and we have $\sqrt{n}(U_n - \theta) \xrightarrow{p} 0$. Indeed, if $\sigma_1^2 = 0$ and $\sigma_2^2 > 0$, then scaling U_n by a factor of n results in a non-Gaussian asymptotic distribution. To present this result, we write χ_1^2 for the chi-square distribution with one degree of freedom and define A_k to be the operator that acts via $g(\cdot) \mapsto E[(k_2(\cdot, Z_1) - \theta)g(Z_1)]$ on square-integrable functions g (that is, $E[g(Z_1)^2] < \infty$).

Theorem 2.3. *If the kernel k of the statistic U_n from (5) has variance $\sigma_m^2 < \infty$ and $\sigma_1^2 = 0 < \sigma_2^2$, then*

$$n(U_n - \theta) \xrightarrow{d} \binom{m}{2} \sum_{i=1}^{\infty} \lambda_i (\chi_{1i}^2 - 1)$$

where $\chi_{11}^2, \chi_{12}^2, \ldots$ are i.i.d. χ_1^2 random variables, and the λ_i's are the eigenvalues, taken with multiplicity, associated with a system of orthonormal eigenfunctions of the operator A_k.

We will use Theorem 2.3 in Section 4 to find the asymptotic distribution of t^* under the null hypothesis of independence.

3. Degeneracy of the sign covariance

Let $Z_i = (X_i, Y_i)$ for $i = 1, 2, \ldots, n$ be an i.i.d. sequence comprising copies of a random vector (X, Y) with values in $\mathbb{R}^2$. Let t^* be the (empirical) Bergsma-Dassios sign covariance for this sample. We begin our study of the asymptotic properties of the U-statistic t^* by studying its degeneracy. Our first observation is that t^* is degenerate when X and Y are independent, denoted $X \perp\!\!\!\perp Y$. Next, in a particular setting that has (X, Y) continuously distributed, we are able to show that t^* is degenerate only if $X \perp\!\!\!\perp Y$.

The statistic t^* has the kernel h from (3), which has $m = 4$ arguments. Specializing the definitions from (6) and (7) to the present setting, we may define functions $h_1, \ldots, h_4$ with variances $\sigma_1^2, \ldots, \sigma_4^2$. The kernel h is a bounded function and thus $\sigma_4^2 < \infty$. Hence, Theorem 2.2 applies and yields the following result.

Corollary 3.1. *As $n \to \infty$, the sign covariance converges to a normal limit, namely,*

$$\sqrt{n}(t^* - \tau^*) \xrightarrow{d} N(0, 16\sigma_1^2).$$

The result just stated provides a non-trivial distributional approximation to t^* only if $\sigma_1^2 > 0$. The following lemma observes that this fails to be the case under the null hypothesis of independence, under which t^* is a degenerate U-statistic. The proof of the lemma as well as the proofs of all other results in this section are deferred to Appendix A.

Lemma 3.1. *If $X \perp\!\!\!\perp Y$ then $\sigma_1^2 = Var[h_1(X_1, Y_1)] = 0$ so that $h_1(X_1, Y_1)$ is a degenerate random variable.*

According to Lemma 3.1 and Theorem 2.2, if $X \perp\!\!\!\perp Y$ we have $\sqrt{n}t^* \xrightarrow{p} 0$ because $E[t^*] = \tau^* = 0$ under independence [1]. We thus need to appeal to Theorem 2.3 to find a non-degenerate asymptotic distribution for t^* when $X \perp\!\!\!\perp Y$. This is the topic of Section 4.

Remark 3.1. *In the continuous case with $X \perp\!\!\!\perp Y$, it is possible to compute all of the variances $\sigma_1^2, \dots, \sigma_4^2$ exactly. We report these values to be*

$$\sigma_1^2 = 0, \quad \sigma_2^2 = \frac{1}{225}, \quad \sigma_3^2 = \frac{8}{225}, \quad \sigma_4^2 = \frac{50}{225}.$$

The fact that $\sigma_1^2 = 0$ was shown in generality in Lemma 3.1. The value of σ_2^2 can be computed as the sum of the squared eigenvalues of h_2 which are derived in the proof of Theorem 4.4; in particular, we have that $\sigma_2^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{6^2}{\pi^8 i^4 j^4} = \frac{1}{225}$. Finally, σ_3^2 and σ_4^2 can be computed explicitly from the representation of h in Lemma 2.1, this computation is trivial for σ_4^2 but quite lengthy for σ_3^2 and thus is omitted.

Next, we turn our attention to the case that $X \not\perp\!\!\!\perp Y$ and (X, Y) are generated from a continuous distribution on $\mathbb{R}^2$. In this case we find t^* to be non-degenerate.

Theorem 3.2. *Suppose (X, Y) has a bivariate continuous distribution with a continuous density function f with support $\overline{f^{-1}((0,\infty))} = [a, b] \times [c, d]$, where $-\infty \le a < b \le \infty$ and $-\infty \le c < d \le \infty$. If X and Y are dependent, then $\sigma_1^2 = \mathit{Var}[h_1(X_1, Y_1)] > 0$.*

In the setting of Theorem 3.2, we thus have that $t^* = N(\tau^*, 16\sigma_1^2/n) + o_p(n^{-1/2})$.

Example 2. *To gain intuition for the magnitude of the asymptotic variance $16\sigma_1^2$, we use Monte Carlo integration to compute $16\sigma_1^2$ in the case that (X, Y) follow a bivariate normal distribution. Since σ_1^2 is an even function of the correlation ρ of a bivariate normal distribution, we consider $\rho \in [0, 1]$. In particular, we perform this computation letting ρ take on 20, evenly spaced, values between 0 and 1. The results of this computation are shown in Figure 2. The figure shows that the asymptotic variance $16\sigma_1^2$ gradually increases with larger values of ρ, peaking at $16\sigma_1^2 \approx 0.14$ when $\rho \approx 0.74$, and then decreases to 0 as the correlation further approaches 1. Note that a value of $\sigma_1^2 = 0$ when $\rho = 1$ does not contradict Theorem 3.2 as, in this case, the joint distribution of (X, Y) is not continuous. The shape of the curve in Figure 2 can be partially explained by the fact that $\sigma_1^2 \le \sigma_4^2 = \mathit{Var}(h(Z_1, \dots, Z_4)) = (\tau^* + 1/3)(1 - (\tau^* + 1/3))$. For instance, note that $(\tau^* + 1/3)(1 - (\tau^* + 1/3))$ equals 0 when $\tau^* = 2/3$ (in which case the correlation can be seen to be 1 or -1), and is maximized at $\tau^* = 1/6 \approx .167$ which corresponds a correlation of approximately 0.7 (see Figure 1).*

4. Asymptotics under the null hypothesis of independence

As in the previous section, let t^* be the empirical sign covariance for an i.i.d. sample (X_i, Y_i), $i = 1, 2, \dots, n$, with values in $\mathbb{R}^2$. Throughout this section, we assume the (X_i, Y_i) to be independent copies of a random vector (X, Y) with $X \perp\!\!\!\perp Y$, so that t^* is degenerate (Lemma 3.1). We thus need to appeal to

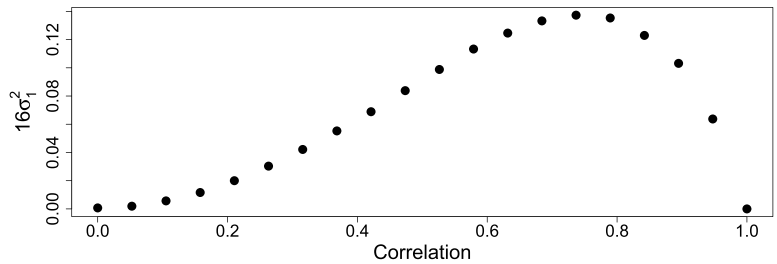


FIG 2. *Monte Carlo approximations to the values of* $16\sigma_1^2$ *for bivariate normal distribution with different correlations.*

Theorem 2.3 to find a non-degenerate asymptotic distribution for t^*. Since $E[t^*] = \tau^* = 0$ under independence, we are led to the problem of determining the eigenvalues of the operator $A_h : g(\cdot) \mapsto E[h_2(\cdot, Z_1)g(Z_1)]$.

A key observation is that under independence A_h is a tensor product of operators because the function h_2 admits the following factorization, which along with all other results in this section is proved in Appendix B.

Lemma 4.1. *If* $X \perp\!\!\!\perp Y$ *then*

$$h_2((x_1, y_1), (x_2, y_2)) = \frac{2}{3}\, g_X(x_1, x_2)\, g_Y(y_1, y_2),$$

where $g_X(x_1, x_2) = E[a(x_1, x_2, X_3, X_4)]$ *and* $g_Y(y_1, y_2) = E[a(y_1, y_2, Y_3, Y_4)]$.

The function g_X (and similarly g_Y) takes the form

$$\begin{aligned} g_X(x_1, x_2) = {} & P(x_1, X_3 < x_2, X_4) + P(x_1, X_3 > x_2, X_4) \\ & - P(x_1, x_2 < X_3, X_4) - P(x_1, x_2 > X_3, X_4). \end{aligned} \tag{8}$$

By Lemma 4.1, $A_h = A_{g_X} \otimes A_{g_Y}$ and thus the spectrum of A_h is the product of the spectra of A_{g_X} and A_{g_Y}. We record the general version of this fact in the next lemma. Here, eigenvalues are always repeated according to their multiplicity, and we let $\mathbb{N}_+ = \{1, 2, \ldots\}$.

Lemma 4.2. *Let* g_1 *and* g_2 *be symmetric real-valued functions with* $E[g_1(X_1, X_2)] = E[g_2(Y_1, Y_2)] = 0$ *and* $E[g_1(X_1, X_2)^2], E[g_2(Y_1, Y_2)^2] < \infty$. *For* $i = 1, 2$, *let* $\lambda_{i,j}$, $j \in \mathbb{N}_+$, *be the nonzero eigenvalues of* A_{g_i}. *Then the products* $\lambda_{1,j_1}\lambda_{2,j_2}$, $j_1, j_2 \in \mathbb{N}_+$, *are the nonzero eigenvalues of* A_k *for* $k((x_1, y_1), (x_2, y_2)) := g_1(x_1, x_2)g_2(y_1, y_2)$.

In the sequel, we use the factorization results from Lemmas 4.1 and 4.2 to obtain the asymptotic distribution of t^* when X and Y are continuous (Section 4.1), and when X and Y are discrete with finite support (Section 4.2). A straightforward extension covers the mixed continuous and discrete case (Section 4.2).

4.1. Continuous variables

Suppose now that $X \perp\!\!\!\perp Y$ with X and Y following continuous marginal distributions. Since $h((X_1, Y_1), \ldots, (X_4, Y_4))$ depends only on the joint ranks of $(X_1, Y_1), \ldots, (X_4, Y_4)$, it follows that τ^* (and t^*) are invariant to monotonically increasing transformations of the marginals of (X, Y). As such we may, and will, assume that X and Y are i.i.d. Uniform$(0, 1)$. Then (X, Y) is uniform on the unit square $(0, 1) \times (0, 1)$. In this case the factorization described in Lemma 4.1 has a particularly nice form.

Lemma 4.3. *If* $X, Y \overset{i.i.d.}{\sim}$ *Uniform*$(0, 1)$*, then for* $(x_1, y_1), (x_2, y_2) \in (0, 1)^2$*,*

$$h_2((x_1, y_1), (x_2, y_2)) = 6\, c(x_1, x_2)\, c(y_1, y_2)$$

where

$$c(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1 \vee x_2 + \frac{1}{3}$$

and $x_1 \vee x_2 := \max\{x_1, x_2\}$*.*

Somewhat surprisingly, the function c corresponds to the kernel of the well studied Cramér-von Mises statistic. Leveraging that the eigenvalues of A_c are already known, we are now able to derive the asymptotic distribution of t^*.

Theorem 4.4. *If* X *and* Y *are independent continuous random variables, then*

$$nt^* \xrightarrow{d} \frac{36}{\pi^4} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{i^2 j^2} (\chi^2_{1,ij} - 1)$$

where $\{\chi^2_{1,ij} : i, j \in \mathbb{N}_+\}$ *is a collection of i.i.d.* χ^2_1 *random variables.*

Remarkably the asymptotic distribution just given is simply a scale multiple of the asymptotic distribution of the U-statistic for Hoeffding's D where

$$D = \iint (F_{X,Y}(x, y) - F_X(x)F_Y(y))^2 \, \mathrm{d}F_{X,Y}(x, y);$$

see [6]. When (X, Y) has a continuous joint distribution, it is readily seen that $D = 0$ if and only if $X \perp\!\!\!\perp Y$. However, this may fail in non-continuous cases.

4.2. Discrete variables

We now treat the case where X and Y are independent discrete random variables with finite supports. Unlike in the continuous case, the asymptotic distribution of t^* then depends on how X and Y distribute their probability mass marginally. In practical applications these marginal probabilities must be estimated before using our limit theorem.

In order to present the result, we associate a matrix to a discrete random variable as follows. Let U be a random variable with finite support $\{u_1, \ldots, u_r\}$, cumulative distribution function F, and probability mass function p. We then define R^U to be the $r \times r$ symmetric matrix whose (i,j)-th entry is

$$R^U_{ij} = \sqrt{p(u_i)p(u_j)}\Bigg\{\Big[(F(u_i \wedge u_j) - p(u_i \wedge u_j))^2 + (1 - F(u_i \vee u_j))^2\Big] - I(u_i \neq u_j)\Bigg[F(u_i \wedge u_j)(1 - F(u_i \wedge u_j)) + \sum_{u_i \wedge u_j < u_\ell < u_i \vee u_j} p(u_\ell)(1 - F(u_\ell))\Bigg]\Bigg\}. \quad (9)$$

Theorem 4.5. *Let X and Y be independent discrete random variables with finite supports of size r and s, respectively. Let $\lambda_1^X, \dots, \lambda_r^X$ be the eigenvalues of R^X, and let $\lambda_1^Y, \dots, \lambda_s^Y$ be the eigenvalues of R^Y. Then*

$$nt^* \xrightarrow{d} 4 \sum_{i=1}^{r} \sum_{j=1}^{s} \lambda_i^X \lambda_j^Y (\chi^2_{1,ij} - 1)$$

where $\{\chi^2_{1,ij} : i \leq r,\ j \leq s\}$ is a collection of rs i.i.d. χ^2_1 random variables.

In the special case that X and Y are Bernoulli random variables, the asymptotic distribution can be presented in simple form.

Example 3. *If $X \sim$ Bernoulli(p) for $p \in (0,1)$, then*

$$R^X = \begin{pmatrix} p^2(1-p) & -(p(1-p))^{3/2} \\ -(p(1-p))^{3/2} & p(1-p)^2 \end{pmatrix}$$

has rank one and its nonzero eigenvalue is $p(1-p)$. It follows that if Y is a second independent random variable with $Y \sim$ Bernoulli(q) for $q \in (0,1)$, then

$$nt^* \xrightarrow{d} 4pq(1-p)(1-q)(\chi^2_1 - 1).$$

So, t^* can be centered and scaled to become asymptotically chi-square.

Example 4. *For a ternary random variable X with $P(X=1) = p_1$, $P(X=2) = p_2$ and $P(X=3) = p_3 = 1 - p_1 - p_2$, we have*

$$R^X = \begin{pmatrix} p_1(1-p_1)^2 & -\sqrt{p_1p_2}\left[p_1(1-p_1) - p_3^2\right] & -\sqrt{p_1p_3}\left[p_3(1-p_3) + p_1p_2\right] \\ \cdot & p_2\left(p_1^2 + p_3^2\right) & -\sqrt{p_2p_3}\left[p_3(1-p_3) - p_1^2\right] \\ \cdot & \cdot & p_3(1-p_3)^2 \end{pmatrix},$$

where we show only the upper half of the symmetric matrix. No simple formula seems to be available to determine the eigenvalues of R^X in this case, but the eigenvalues can readily be computed numerically for any (possibly estimated) values of p_1 and p_2.

Finally, if X is discrete with finite support and Y is continuous, then a simple extension of Theorems 4.4 and 4.5 gives the following result.

Corollary 4.1. *Let X and Y be independent random variables, where X has finite support of size r and Y is continuous. Let $\lambda_1, \dots, \lambda_r$ be the eigenvalues of R^X. Then*

$$nt^* \xrightarrow{d} \frac{12}{\pi^2} \sum_{i=1}^{r} \sum_{j=1}^{\infty} \frac{\lambda_i}{j^2} (\chi^2_{1,ij} - 1)$$

where $\{\chi^2_{1,ij} : i \leq r,\ j \in \mathbb{N}_+\}$ is a collection of i.i.d. χ^2_1 random variables.

5. Simulations

The results from Section 4 can be used to form asymptotic tests of independence, and we now explore which sample sizes are needed for the asymptotic approximations to be accurate. As a test based on t^* has asymptotic power against all alternatives to independence, it is also of interest to make comparisons against other tests known to be (most) powerful for particular settings and alternatives. Finally, we demonstrate how the results of Section 3 can be used for sample size computations. Code for performing asymptotic tests of independence has been incorporated in the TauStar[1] R package available on CRAN, the Comprehensive R Archive Network [8, 14].

5.1. Empirical convergence to the asymptotic distribution

Let t^* be computed from a sample of size n drawn from the joint distribution of a bivariate random vector (X, Y) with $X \perp\!\!\!\perp Y$. Since t^* only depends on ranks, its distribution does not change when applying monotonically increasing marginal transformations to X and Y. When X and Y both have continuous distributions, we may thus transform their distributions to $N(0, 1)$ without changing the distribution of t^*. When one or both of X and Y are discrete however, the distribution of t^* depends on how X and Y distribute their probability mass making it impossible to provide an exhaustive empirical study of convergence properties. Instead we will consider selected examples. Specifically, we consider the following cases:

- (i) The continuous case with $X, Y \sim N(0, 1)$.
- (ii) A discrete case with $P(X = i) = 1/10$ for $1 \leq i \leq 10$, and $P(Y = i) \propto 2^{-i}$ for $1 \leq i \leq 12$.

[1] See https://cran.r-project.org/web/packages/TauStar/index.html

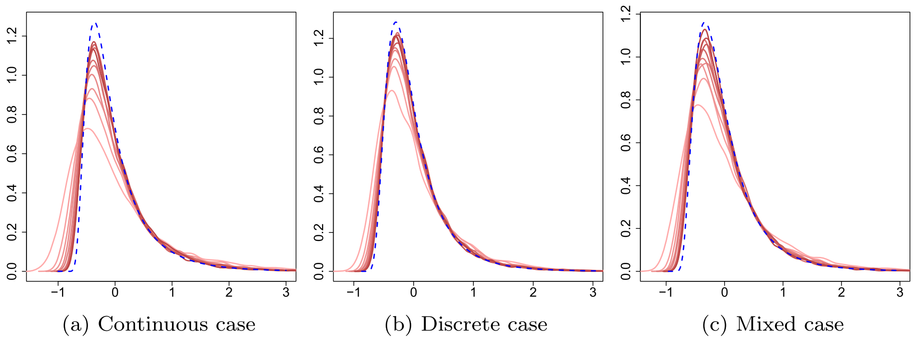


FIG 3. *Kernel density estimates from 10,000 simulated values of* nt^* *at sample sizes* $n \in \{10, 15, 20, 25, 30, 40, 50, 60, 70, 80\}$*; smaller values of* n *are shown in lighter color. The plots also show the density of the asymptotic distributions from Section 4 in dashed blue line.*

(iii) A mixed case with $X \sim N(0,1)$ and $P(Y = i) = 1/5$ for $1 \leq i \leq 5$.

In each setting we compute, for different sample sizes n, a kernel density estimate for the distribution of t^* and plot it alongside the asymptotic density. The resulting plots are shown in Figure 3, which demonstrates that the asymptotic and finite-sample distributions are in close agreement already when $n = 80$. While we present only one example each for the discrete and mixed cases we found similar results when simulating with many other choices of distributions.

Remark 5.1. *Computing the asymptotic densities shown in Figure 3 is non-trivial and requires the numerical inversion of the characteristic function for the asymptotic distributions. To perform this numerical inversion we use the techniques described in Section 7 of Blum, Kiefer, and Rosenblatt [2]; these computations are done automatically in the aforementioned TauStar package for R.*

5.2. Power comparisons

We explore the power of an asymptotic test based on t^* in six cases:

(i) First, we take (X, Y) as bivariate normal with correlation $\rho \in \{0, .1, .2, \ldots, 1\}$; the distribution of t^* then does not depend on the means and variances which may thus be set to zero and one, respectively. We compare the test based on t^* to the two-sided test based on the standard Pearson correlation $\widehat{\rho}$. We implement the latter test using the fact that $\widehat{\rho}\sqrt{(n-2)/(1-\widehat{\rho}^2)}$ has a t-distribution with $n - 2$ degrees of freedom.

(ii) Next, we consider three discrete cases all of which have (X, Y) taking values in the grid $\{1, 2, \ldots, 5\}^2$. In each of these cases we compare our test to the chi-square test of independence.

   (a) In the first discrete case, (X, Y) follows a mixture between the uniform distribution on $\{(1, 1), (2, 2), \ldots, (5, 5)\}$ and the uniform distribution

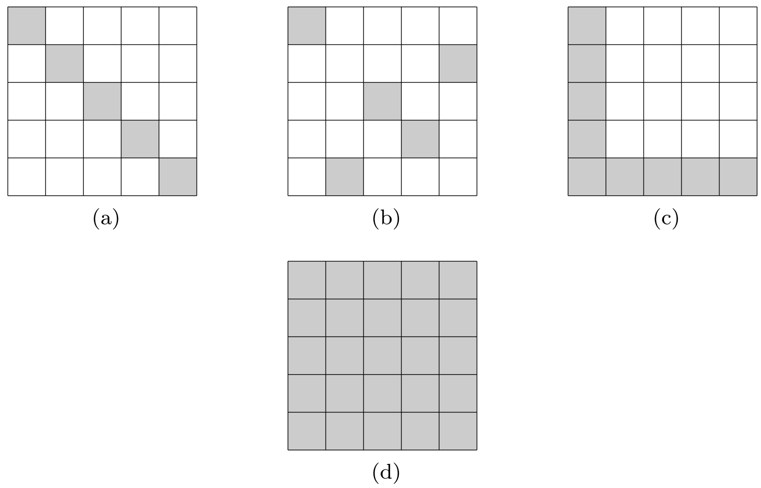


Fig 4. *Visualization of where probability mass is placed for different discrete distributions on* $\{1, \ldots, 5\}^2$*. In each case the distribution is uniform over the gray squares, and zero probability is assigned to the white squares.*

on $\{1, \ldots, 5\}^2$, illustrated in Figures 4a and 4d, respectively. We let the mixture weight p for the former component range through the set $\{0, .1, \ldots, 1\}$.

(b) The second discrete case is analogous but a mixture between the distributions from Figures 4b and 4d.

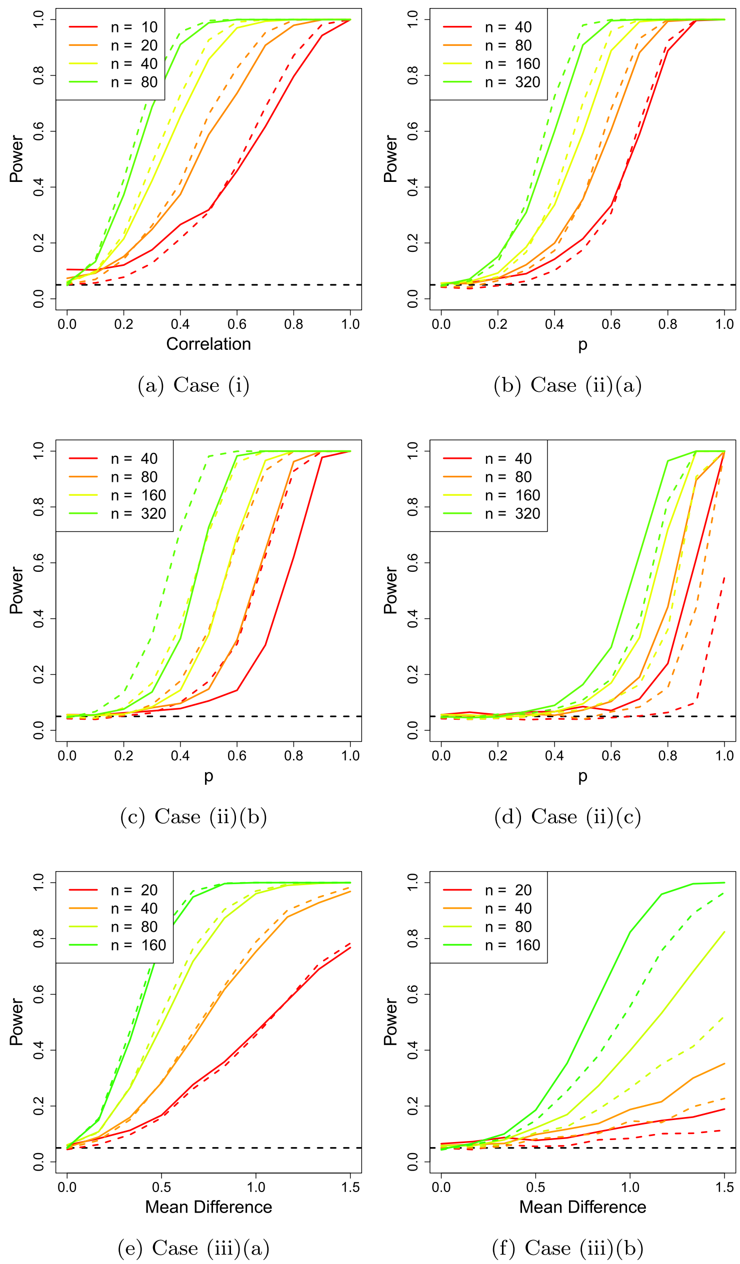
(c) The third discrete case is as the previous two but mixes the distributions from Figures 4c and 4d.

(iii) Finally, we experiment with two mixed cases in which the distribution of X is discrete and the conditional distribution of Y given $X = x$ is a normal distribution $N(\mu_x, 1)$.

(a) The first mixed case has $X \sim$ Bernoulli(.3), $\mu_x = 0$ when $x = 1$, and $\mu_x = \mu$ when $x = 0$. Here, we let the mean difference μ range through the set $\{0, 1/6, 2/6, \ldots, 9/6\}$, and each setting we compare against the two-sample t-test.

(b) In the second case $X \sim$ Uniform($\{1, \ldots, 6\}$), and Y has conditional mean μ_x is zero when x is odd and equal to $\mu \in \{0, 1/6, 2/6, \ldots, 9/6\}$ when x is even. Here, we compare against a bootstrapped permutation test using the distance covariance statistic of [12], known to be consistent for independence, using the Energy R package [9].

The simulation results are presented in Figure 5. Surprisingly, the t^* test has competitive power in cases (i) and (iii)(a) where the alternative tests are known to be most powerful given the distributional assumption of normality. For the jointly discrete cases, we observe that the chi-square test of independence has



(a) Case (i) (b) Case (ii)(a)

(c) Case (ii)(b) (d) Case (ii)(c)

(e) Case (iii)(a) (f) Case (iii)(b)

FIG 5. *Simulated power of the* t^* *asymptotic test (in solid line) and the power of the competing test (in dashed line). In each case we use a level of 0.05, which is displayed as a horizontal dashed black line.*

essentially equal power in case (ii)(a), significantly higher power in case (ii)(b), and significantly lower power in case (ii)(c). The lack of power in case (ii)(b) is not surprising as the t^* statistic is ordinal in nature and the dependence in the distribution from case (ii)(b) was designed to be non-ordinal. The ordinal nature of t^* also explains the significant gains in case (ii)(c). Hence, it would seem that the t^* test for jointly discrete data can offer substantial improvements in power over the chi-square test if an ordinal dependence relationship is suspected in the data. Finally, case (iii)(b) suggests that there are cases in which t^* may provide higher power than the distance covariance.

5.3. Sample size calculations

Focusing on the continuous case, consider an asymptotic level α test of the null hypothesis of $\tau^* = 0$ (i.e., independence) that compares the statistic t^* to a critical value c_α derived from the asymptotic distribution from Theorem 4.4. Suppose we would like to determine the minimum sample size n_β needed for a power of at least β under an alternative that has the two considered variables X and Y dependent, so that $\tau^* > 0$. If the sample is drawn from a joint distribution for (X, Y) that satisfies the conditions of Theorem 3.2, and if σ_1^2 is known to be no larger than the quantity $\bar{\sigma}_1^2$, then Corollary 3.1 implies that for any $x \le \tau^*$,

$$
\begin{aligned}
P(t^* \le x) &= P\left(\sqrt{n}(t^* - \tau^*) \le \sqrt{n}(x - \tau^*)\right) \\
&\approx P\left(N(0, 16\,\sigma_1^2) \le \sqrt{n}(x - \tau^*)\right) \\
&\le P\left(N\left(\tau^*, 16\,\bar{\sigma}_1^2/n\right) \le x\right).
\end{aligned}
$$

This result can be used to find an asymptotically valid upper bound $\bar{n}_\beta$ on n_β, by letting $\bar{n}_\beta$ be the smallest positive integer such that $c_\alpha/\bar{n}_\beta \le \tau^*$ and $P(N(\tau^*, 16\,\bar{\sigma}_1^2/\bar{n}_\beta) \le c_\alpha/\bar{n}_\beta) \le 1 - \beta$. Finding this number $\bar{n}_\beta$ can be accomplished in an iterative fashion.

The remaining difficulty in such an asymptotic sample size calculation is finding a suitable upper bound $\bar{\sigma}_1^2$ for the unknown variance σ_1^2. A crude but universally valid upper bound for σ_1^2 can be obtained from Lemma 2.1, which implies that $h_1(X, Y)$ takes values in the interval $[-1/3, 2/3]$ and thus $\sigma_1^2 \le 1/4$. When (X, Y) is bivariate normal, an approximately valid upper bound of σ_1^2 is given by $\sigma_1^2 \le 0.14/16 = 0.00875$ (see Example 2). Figure 6 plots the upper bound for the minimum sample size needed to achieve various powers when bounding σ_1^2 by $1/4$ and 0.00875 respectively and sampling from a bivariate normal distribution with correlation 0.6. From the figure, we see that the $1/4$ bound leads to very conservative sample sizes while the 0.00875 bound results in values that much more closely adhere to the empirical truth. In general, overestimation of σ_1^2 is advisable as small values of $\bar{\sigma}_1^2$ may lead to consideration of sample sizes that are too small for asymptotic approximations to be reflective of the actual finite-sample behavior of the test.

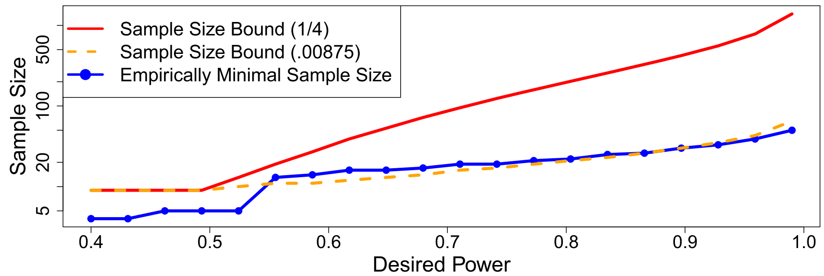


FIG 6. *Minimum sample size (n_β) needed to a achieve a desired power β at level 0.05. Simulations for bivariate normal data with correlation 0.6 were used to compute an estimate of n_β (blue line with dots). These are compared to two asymptotic upper bounds for n_β using the bound $\sigma_1^2 \le 1/4$ (red line) and the bound $\sigma_1^2 \le 0.00875$ (dashed orange line). The sample size is presented with a log scaling.*

6. Discussion

The sign covariance τ^* of [1] has the intriguing property of being zero if and only if the considered pair of random variables is independent, assuming that the random variables follow a distribution that is continuous, discrete or a mixture of such distributions. Under these mild conditions, testing the hypothesis that $\tau^* = 0$ thus allows one to consistently assess (in-)dependence. With the aim of simplifying the implementation of such independence tests, we have given a comprehensive study of the asymptotic properties of t^*, the natural U-statistic for τ^*. The asymptotic distribution of t^*, especially as described in Section 4.1, is seen to be connected in interesting ways to the asymptotic distribution of Hoeffding's D, and the Cramér-von Mises statistic.

After the completion of our manuscript we found that some asymptotic properties of t^* have been considered by Dhar *et al.* [3]. Indeed, their work provides an excellent complement to the results presented here as it considers the large-sample behavior of t^* under local contiguous alternatives while ours provides explicit descriptions of asymptotic distributions under independence and illuminates interesting large-sample phenomena under fixed alternatives.

While we have a complete understanding of the asymptotics of t^* under fairly weak distributional assumptions—we covered continuous and discrete cases, it remains to be seen if the large-sample distribution of t^* can be obtained without any such assumptions. However, as noted above, it is also not yet known if the property that $\tau^* = 0$ only under independence holds for distributions that are not (absolutely) continuous, discrete or a mixture of two such distributions.

Appendix A: Proofs for Section 3

Proof of Lemma 3.1. We show that $h_1(x_1, y_1) = 0$ for any (x_1, y_1) in the support of (X, Y). Since X and Y are independent and $X_1, \dots, X_4$ as well as

$Y_1, \dots, Y_4$ are i.i.d. random variables, we have

$$\begin{aligned} &h_1((x_1, y_1)) \\ &\quad = \frac{1}{4!} \sum_{\pi \in S_4} E\left[a(X_{\pi(1,2,3,4)}) \mid X_1 = x_1\right] E\left[a(Y_{\pi(1,2,3,4)}) \mid Y_1 = y_1\right] \\ &\quad = \frac{1}{4!} \sum_{\pi \in S_4} E\left[a(X_1, X_2, X_3, X_4) \mid X_{\pi(1)} = x_1\right] E\left[a(Y_1, Y_2, Y_3, Y_4) \mid Y_{\pi(1)} = y_1\right]. \end{aligned}$$

Thus it suffices to show that $g_X^{(j)}(x_1) := E\left[a(X_1, X_2, X_3, X_4) \mid X_j = x_1\right] = 0$, for $j = 1, 2, 3, 4$. For $j = 1$, we have

$$\begin{aligned} g_X^{(1)}(x_1) &= P(x_1, X_3 < X_2, X_4) + P(x_1, X_3 > X_2, X_4) \\ &\quad - P(x_1, X_2 < X_3, X_4) - P(x_1, X_2 > X_3, X_4) \\ &= 0 \end{aligned}$$

because X_2, X_3, X_4 are i.i.d. and thus exchangeable. Analogous arguments show that all other $g_X^{(j)}(x_1)$ are zero. □

Proof of Theorem 3.2. Let F be the, by assumption, continuously differentiable joint distribution function of (X, Y), and let F_X and F_Y be the two marginal distribution functions. Since h is invariant to monotonically increasing transformations of its coordinates, we may assume without loss of generality that we have applied F_X and F_Y to (X, Y) coordinate-wise, so that X and Y are Uniform(0,1) marginally. Moreover, since we assumed that (X, Y) had support $[a, b] \times [c, d]$ for $a < b,\ c < d$, it follows that (X, Y) has support $[0, 1]^2$ after the transformation. The main idea of the proof is to show

$$\left. \frac{\partial^2}{\partial y_1 \partial x_1} h_1(x_1, y_1) \right|_{(x^*, y^*)} \neq 0 \text{ for some } (x^*, y^*) \in (0, 1)^2. \tag{10}$$

A continuity argument and (10) then imply that $h_1(x, y)$ is a non-constant function on a set of non-zero probability and thus $h_1(X, Y)$ is non-degenerate.

Note that since $X, Y \sim \text{Uniform}(0, 1)$ marginally we have that $F_X(x) = x$ and $F_Y(y) = y$ for all $x, y \in [0, 1]$. Consequently, the marginal densities of X and Y, $f_X(x) := \frac{\partial}{\partial x} F_X(x)$ and $f_Y(y) := \frac{\partial}{\partial y} F_Y(y)$, equal 1 on $[0, 1]$. We write f for the probability density function of (X, Y), which is assumed continuous, and we denote the conditional distribution function of X given $Y = y$ by $F_{X|y}(x)$ and denote the conditional distribution function of Y given $X = x$ by $F_{Y|x}(y)$. In Lemma A.1 below, we find that

$$\frac{\partial^2 h_1(x_1, y_1)}{\partial y_1 \partial x_1} = 6G(x_1, y_1)[2f(x_1, y_1) + 1] + 6[F_{Y|x_1}(y_1) - y_1][F_{X|y_1}(x_1) - x_1], \tag{11}$$

where $G(x_1, y_1) = F(x_1, y_1) - F_X(x_1) F_Y(y_1) = F(x_1, y_1) - x_1 y_1$. We proceed to show how to derive (10) from (11).

Since $\frac{\partial}{\partial x}F(x,y) = F_{Y|x}(y)f_X(x) = F_{Y|x}(y)$ and similarly $\frac{\partial}{\partial y}F(x,y) = F_{X|y}(x)$, we have $\frac{\partial}{\partial x_1}G(x_1,y_1) = F_{Y|x_1}(y_1) - y_1$ and $\frac{\partial}{\partial y_1}G(x_1,y_1) = F_{X|y_1}(x_1) - x_1$. Thus

$$\frac{\partial^2 \ h_1(x_1,y_1)}{\partial y_1 \partial x_1} = 6G(x_1,y_1)[2f(x_1,y_1)+1] + 6\left[\frac{\partial}{\partial x_1}G(x_1,y_1)\right]\left[\frac{\partial}{\partial y_1}G(x_1,y_1)\right].$$

Now, G is continuous because F is, and thus the compactness of $[0,1]^2$ yields that G attains its extrema on $[0,1]^2$. In other words, there exist $z_m = (x_m, y_m)$, $z_M = (x_M, y_M) \in [0,1]^2$ such that $G(z_m) = \inf_{(x,y)\in[0,1]^2} G(x,y)$, $G(z_M) = \sup_{(x,y)\in[0,1]^2} G(x,y)$. Since X and Y are dependent we must have that either $G(z_M) > 0$ or $G(z_m) < 0$. Without loss of generality assume that $G(z_M) > 0$.

The support of (X,Y) being equal to $[0,1]^2$, we have that $G(x,y) = 0$ for all (x,y) on the boundary of $[0,1]^2$. Hence, $z_M = (x_M, y_M)$ lies in the interior of $[0,1]^2$ and as a local (global) maximum of G, it satisfies

$$\frac{\partial}{\partial x_1}G(x_1,y_1)\bigg|_{(x_M,y_M)} = \frac{\partial}{\partial y_1}G(x_1,y_1)\bigg|_{(x_M,y_M)} = 0.$$

We deduce that (10) because

$$\frac{\partial^2}{\partial y_1 \partial x_1}h_1(x_1,y_1)\bigg|_{(x_M,y_M)} = 6G(x_M,y_M)[2f(x_M,y_M)+1] > 0.$$

(If instead we had assumed that $G(z_m) < 0$ then the same arguments would hold and the above inequality would be < 0 instead of > 0.)

Finally, $\frac{\partial^2}{\partial y_1 \partial x_1}h_1(x_1,y_1)$ is easily seen to be continuous and thus we have that $\frac{\partial^2}{\partial y_1 \partial x_1}h_1(x_1,y_1) > 0$ in an open neighborhood U of z_M. Since the support of $f(x,y)$ is all of $[0,1]^2$, that is $\overline{f^{-1}((0,\infty))} = [0,1]^2$, it follows that $U \cap f^{-1}((0,\infty))$ is a non-empty open set and thus the claim of the theorem follows. □

Lemma A.1. *Let (X,Y) have joint density f and joint distribution function F. Let F_X and F_Y be the marginal distribution functions, and let $F_{X|y}$ and $F_{Y|x}$ be the conditional distribution functions of X given $Y=y$ and Y given $X=x$, respectively. If $X,Y \sim$ Uniform$(0,1)$ marginally, then*

$$\frac{\partial^2 \ h_1(x_1,y_1)}{\partial y_1 \partial x_1} = 6G(x_1,y_1)[2f(x_1,y_1)+1] + 6[F_{Y|x_1}(y_1) - y_1][F_{X|y_1}(x_1) - x_1],$$

where $G(x_1,y_1) = F(x_1,y_1) - F_X(x_1)F_Y(y_1) = F(x_1,y_1) - x_1y_1$.

Proof. Let $Z_1 = (X_1,Y_1), \dots, Z_4 = (X_4,Y_4)$ be i.i.d. copies of (X,Y). $Z_1, \dots, Z_4$ are almost surely either concordant or discordant. It follows from Lemma 2.1 that

$$h(Z_1,Z_2,Z_3,Z_4) = I(Z_1,Z_2,Z_3,Z_4 \text{ are concordant}) - 1/3,$$

where $I(\cdot)$ is the indicator function as usual. Let $C(Z_1,\dots,Z_4)$ denote the event that $Z_1,\dots,Z_4$ are concordant. Then

$$\begin{aligned} h_1(x_1,y_1)+1/3 &= P(C(z_1,Z_2,Z_3,Z_4)) \\ &= 3P(x_1 \le X_2 \le X_3, X_4 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad + 3P(X_2 \le x_1 \le X_3, X_4 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad + 3P(X_3, X_4 \le x_1 \le X_2 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad + 3P(X_3, X_4 \le X_2 \le x_1 \text{ and } C(z_1,Z_2,Z_3,Z_4)). \end{aligned}$$

We make the definitions

$$\begin{aligned} P_{bl}(x,y) &:= P(X \le x, Y \le y), \qquad & P_{tl}(x,y) &:= P(X \le x, Y > y), \\ P_{br}(x,y) &:= P(X > x, Y \le y), \qquad & P_{tr}(x,y) &:= P(X > x, Y > y). \end{aligned}$$

As suggested by the notation, $P_{bl}(x,y)$ is the probability of (X,Y) being in the 'bottom left' quadrant when dividing $\mathbb{R}^2$ by the lines $\{x\}\times\mathbb{R}$ and $\mathbb{R}\times\{y\}$, and the notation for the other three probabilities is motivated similarly. Now note that

$$\begin{aligned} &P(x_1 \le X_2 \le X_3, X_4 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad = \int_0^1 \int_{x_1}^1 [P(X_3, X_4 > x \text{ and } Y_3, Y_4 \le \min(y_1, y)) \\ &\qquad + P(X_3, X_4 > x \text{ and } Y_3, Y_4 > \max(y_1,y))] f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ &\quad = \int_0^{y_1} \int_{x_1}^1 \{P_{br}^2(x,y) + P_{tr}^2(x,y_1)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ &\qquad + \int_{y_1}^1 \int_{x_1}^1 \{P_{br}^2(x,y_1) + P_{tr}^2(x,y)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y. \end{aligned}$$

Similarly, we have

$$\begin{aligned} &P(X_2 \le x_1 \le X_3, X_4 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad = \int_0^{y_1} \int_0^{x_1} \{P_{br}^2(x_1,y) + P_{tr}^2(x_1,y_1)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ &\qquad + \int_{y_1}^1 \int_0^{x_1} \{P_{br}^2(x_1,y_1) + P_{tr}^2(x_1,y)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y, \\ &P(X_3, X_4 \le x_1 \le X_2 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad = \int_0^{y_1} \int_{x_1}^1 \{P_{bl}^2(x_1,y) + P_{tl}^2(x_1,y_1)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ &\qquad + \int_{y_1}^1 \int_{x_1}^1 \{P_{bl}^2(x_1,y_1) + P_{tl}^2(x_1,y)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y, \end{aligned}$$

and

$$\begin{aligned} &P(X_3, X_4 \le X_2 \le x_1 \text{ and } C(z_1,Z_2,Z_3,Z_4)) \\ &\quad = \int_0^{y_1} \int_0^{x_1} \{P_{bl}^2(x,y) + P_{tl}^2(x,y_1)\} f(x,y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

$$+\int_{y_1}^{1}\int_{0}^{x_1}\{P_{bl}^2(x,y_1)+P_{tl}^2(x,y)\}f(x,y)\ \mathrm{d}x\ \mathrm{d}y.$$

Now, a straightforward but lengthy computation shows that

$$\begin{aligned}
&\frac{\partial^2}{\partial y_1 \partial x_1}(\frac{1}{3}h_1(x_1,y_1)+\frac{1}{9}) \\
&\quad=\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{bl}^2(x_1,y_1)\right\}P_{tr}(x_1,y_1)+\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{tl}^2(x_1,y_1)\right\}P_{br}(x_1,y_1) \\
&\qquad+\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{br}^2(x_1,y_1)\right\}P_{tl}(x_1,y_1)+\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{tr}^2(x_1,y_1)\right\}P_{bl}(x_1,y_1).
\end{aligned} \tag{12}$$

In terms of the distribution function, the quadrant probabilities are

$$\begin{aligned}
&P_{bl}(x_1,y_1)=F(x_1,y_1),\\
&P_{tl}(x_1,y_1)=F_X(x_1)-F(x_1,y_1)=x_1-F(x_1,y_1),\\
&P_{br}(x_1,y_1)=F_Y(y_1)-F(x_1,y_1)=y_1-F(x_1,y_1),\ \text{and}\\
&P_{tr}(x_1,y_1)=1-F_X(x_1)-F_Y(y_1)+F(x_1,y_1)=1-x_1-y_1+F(x_1,y_1).
\end{aligned}$$

Using that $\frac{\partial}{\partial x}F(x,y) = F_{Y|x}(y)$, $\frac{\partial}{\partial y}F(x,y) = F_{X|y}(x)$ and $\frac{\partial^2}{\partial y \partial x}F(x,y) = f(x,y)$, we obtain that

$$\begin{aligned}
&\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{bl}^2(x_1,y_1)\right\}P_{tr}(x_1,y_1)\\
&\quad=2P_{tr}(x_1,y_1)\left[P_{bl}(x_1,y_1)\frac{\partial^2}{\partial y_1 \partial x_1}P_{bl}(x_1,y_1)\right.\\
&\qquad\qquad\left.+\left\{\frac{\partial}{\partial x_1}P_{bl}(x_1,y_1)\right\}\left\{\frac{\partial}{\partial y_1}P_{bl}(x_1,y_1)\right\}\right] && (13)\\
&\quad=2P_{tr}(x_1,y_1)\left[P_{bl}(x_1,y_1)f(x_1,y_1)+F_{Y|x_1}(y_1)F_{X|y_1}(x_1)\right]. && (14)
\end{aligned}$$

Similarly

$$\begin{aligned}
&\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{tl}^2(x_1,y_1)\right\}P_{br}(x_1,y_1)\\
&\quad=-2P_{br}(x_1,y_1)\left[P_{tl}(x_1,y_1)f(x_1,y_1)+(1-F_{Y|x_1}(y_1))F_{X|y_1}(x_1)\right], && (15)\\
&\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{br}^2(x_1,y_1)\right\}P_{tl}(x_1,y_1)\\
&\quad=-2P_{tl}(x_1,y_1)[P_{br}(x_1,y_1)f(x_1,y_1)+F_{Y|x_1}(y_1)(1-F_{X|y_1}(x_1))], && (16)
\end{aligned}$$

and

$$\left\{\frac{\partial^2}{\partial y_1 \partial x_1}P_{tr}^2(x_1,y_1)\right\}P_{bl}(x_1,y_1)$$

$$= 2P_{bl}(x_1,y_1)[P_{tr}(x_1,y_1)f(x_1,y_1) + (1-F_{Y|x_1}(y_1))(1-F_{X|y_1}(x_1))]. \quad (17)$$

Combining (12)–(17), we find that

$$\begin{aligned}
&\frac{1}{3}\frac{\partial^2}{\partial y_1 \partial x_1} h(x_1,y_1) \\
&\quad = 4[P_{bl}(x_1,y_1)P_{tr}(x_1,y_1) - P_{tl}(x_1,y_1)P_{br}(x_1,y_1)]f(x_1,y_1) \\
&\qquad + 2[F(x_1,y_1) + F_{Y|x_1}(y_1)F_{X|y_1}(x_1) - x_1F_{Y|x_1}(y_1) - y_1F_{X|y_1}(x_1)] \\
&\quad = 2[F(x_1,y_1) - x_1y_1][2f(x_1,y_1)+1] + 2[F_{Y|x_1}(y_1) - y_1][F_{X|y_1}(x_1) - x_1],
\end{aligned}$$

which gives the claimed formula.

Appendix B: Proofs for Section 4

Proof of Lemma 4.1. First note that

$$\begin{aligned}
&h_2((x_1,y_1),(x_2,y_2)) \\
&= \frac{1}{4!}\sum_{\pi\in S_4} E\big[a(X_{\pi(1,2,3,4)})\,|\,X_1=x_1,X_2=x_2\big]\, E\big[a(Y_{\pi(1,2,3,4)})\,|\,Y_1=y_1,Y_2=y_2\big] \\
&= \frac{1}{4!}\sum_{\pi\in S_4}\Big(E\left[a(X_1,X_2,X_3,X_4)\mid X_{\pi(1)}=x_1,X_{\pi(2)}=x_2\right] \\
&\qquad\qquad \cdot E\left[a(Y_1,Y_2,Y_3,Y_4)\mid Y_{\pi(1)}=y_1,Y_{\pi(2)}=y_2\right]\Big) \\
&=: \frac{1}{4!}\sum_{\pi\in S_4} g_X^{\pi}(x_1,x_2)g_Y^{\pi}(y_1,y_2). \qquad (18)
\end{aligned}$$

The first equality follows from the independence of X and Y and the second equality follows from the fact that $X_1,\dots,X_4$ (and $Y_1,\dots,Y_4$) are i.i.d. random variables.

Next, recall from (8) that

$$\begin{aligned}
g_X(x_1,x_2) &= P(x_1,X_3<x_2,X_4) + P(x_1,X_3>x_2,X_4) \\
&\quad - P(x_1,x_2<X_3,X_4) - P(x_1,x_2>X_3,X_4).
\end{aligned}$$

We claim that

$$g_X^{\pi}(x_1,x_2) = \begin{cases} g_X(x_1,x_2) & \text{if } \pi(1),\pi(2)\in\{1,2\} \text{ or } \pi(1),\pi(2)\in\{3,4\}, \\ -g_X(x_1,x_2) & \text{if } \pi(1),\pi(2)\in\{1,3\} \text{ or } \pi(1),\pi(2)\in\{2,4\}, \\ 0 & \text{otherwise.} \end{cases} \qquad (19)$$

Note that (19) implies that $g_X^{\pi}(x_1,x_2)$ is nonzero for 16 of the 24 permutations $\pi\in S_4$. For a set of 8 of these permutations, $g_X^{\pi}(x_1,x_2) = g_X(x_1,x_2)$, and for the other 8, $g_X^{\pi}(x_1,x_2) = -g_X(x_1,x_2)$. The analogue is true for $g_Y^{\pi}(y_1,y_2)$.

Taking products and summing over the permutations π as in (18) completes the proof of the formula for $h_2((x_1, y_1), (x_2, y_2))$.

It remains to show the claim in (19). Since X_3 and X_4 are i.i.d. random variables, $\pi(1), \pi(2) \in \{1, 2\}$ implies $g_X^\pi(x_1, x_2) = g_X(x_1, x_2)$ or $g_X(x_2, x_1)$. But g_X is symmetric and thus $\pi(1), \pi(2) \in \{1, 2\}$ implies $g_X^\pi(x_1, x_2) = g_X(x_1, x_2)$. Analogously, it follows that $g_X^\pi(x_1, x_2) = g_X(x_1, x_2)$ if $\pi(1), \pi(2) \in \{3, 4\}$ because

$$\begin{aligned} E[a(X_1, X_2, x_1, x_2)] &= P(X_1, x_1 < X_2, x_2) + P(X_1, x_1 > X_2, x_2) \\ &\quad - P(X_1, X_2 < x_1, x_2) - P(X_1, X_2 > x_1, x_2) \\ &= g_X(x_1, x_2). \end{aligned}$$

Now if $\pi(1) = 1$ and $\pi(2) = 3$, then

$$\begin{aligned} g_X^\pi(x_1, x_2) &= E[a(x_1, X_2, x_2, X_4)] \\ &= P(x_1, x_2 < X_2, X_4) + P(x_1, x_2 > X_2, X_4) \\ &\quad - P(x_1, X_2 < x_2, X_4) - P(x_1, X_2 > x_2, X_4) \\ &= -g_X(x_1, x_2). \end{aligned}$$

Similar symmetry arguments thus yield that $g_X^\pi(x_1, x_2) = -g_X(x_1, x_2)$ if $\pi(1)$, $\pi(2) \in \{1, 3\}$ or if $\pi(1), \pi(2) \in \{2, 4\}$.

In the remaining cases, we have $\pi(1), \pi(2) \in \{1, 4\}$ or $\pi(1), \pi(2) \in \{2, 3\}$. If $\pi(1) = 1$ and $\pi(2) = 4$,

$$\begin{aligned} g_X^\pi(x_1, x_2) &= E[a(x_1, X_2, X_3, x_2)] \\ &= P(x_1, X_3 < X_2, x_2) + P(x_1, X_3 > X_2, x_2) \\ &\quad - P(x_1, X_2 < X_3, x_2) - P(x_1, X_2 > X_3, x_2) \\ &= P(x_1, X_3 < X_2, x_2) + P(x_1, X_3 > X_2, x_2) \\ &\quad - P(x_1, X_3 < X_2, x_2) - P(x_1, X_3 > X_2, x_2) \\ &= 0. \end{aligned}$$

Similarly, $g_X^\pi(x_1, x_2) = 0$ if $\pi(1) = 4$ and $\pi(2) = 1$, or if $\pi(1), \pi(2) \in \{2, 3\}$. □

Proof of Lemma 4.2. Let $\varphi_{i,1}, \varphi_{i,2}, \ldots$ be the sequence of orthonormal eigenfunctions associated with the nonzero eigenvalues $\lambda_{i,1}, \lambda_{i,2}, \ldots$ of A_{g_i}, for $i = 1, 2$. Since $X \perp\!\!\!\perp Y$, for each $(j_1, j_2) \in \mathbb{N}_+^2$,

$$\begin{aligned} &E[k((x_1, y_1), (X_2, Y_2))\varphi_{1,j_1}(X_2)\varphi_{2,j_2}(Y_2)] \\ &\quad = E[g_1(x_1, X_2)g_2(y_1, Y_2)\varphi_{1,j_1}(X_2)\varphi_{2,j_2}(Y_2)] \\ &\quad = E[g_1(x_1, X_2)\varphi_{1,j_1}(X_2)]\, E[g_2(y_1, Y_2)\varphi_{2,j_2}(Y_2)] \\ &\quad = \lambda_{1,j_1}\varphi_{1,j_1}(x_1)\lambda_{2,j_2}\varphi_{2,j_2}(y_1). \end{aligned}$$

Therefore, for each $(j_1, j_2) \in \mathbb{N}_+^2$, $\lambda_{1,j_1}\lambda_{2,j_2}$ is an eigenvalue of A_k with the associated eigenfunction $\varphi_{1,j_1}\varphi_{2,j_2}$. Further, $\{\varphi_{1,j_1}\varphi_{2,j_2} : (j_1, j_2) \in \mathbb{N}_+^2\}$ is an orthonormal system, since both $\{\varphi_{1,1}, \varphi_{1,2}, \ldots\}$ and $\{\varphi_{2,1}, \varphi_{2,2}, \ldots\}$ are orthonormal systems, and $X \perp\!\!\!\perp Y$.

Now suppose $\{\gamma_1, \gamma_2, \ldots\}$ is a sequence of all nonzero eigenvalues of A_k with the associated orthonormal sequence of eigenfunctions $\{\psi_1, \psi_2, \ldots\}$. Then

$$\sum_{j=1}^{n} \gamma_j \psi_j((X_1, Y_1), (X_2, Y_2)) \xrightarrow{L^2} k((X_1, Y_1), (X_2, Y_2)).$$

By independence,

$$\begin{aligned}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} \lambda_{1,j_1}^2 \lambda_{2,j_2}^2 &= E[g_1(X_1, X_2)^2]\ E[g_2(Y_1, Y_2)^2] \\ &= E[(g_1(X_1, X_2) g_2(Y_1, Y_2))^2] \\ &= E[k((X_1, Y_1), (X_2, Y_2))^2] \\ &= \sum_{j=1}^{\infty} \gamma_j^2\end{aligned}$$

Therefore, we conclude that, as a multi-set, $\{\lambda_{1,j_1}\lambda_{2,j_2} : (j_1, j_2) \in \mathbb{N}_+^2\}$ contains all nonzero eigenvalues of A_k with the correct multiplicity. □

Proof of Lemma 4.3. Any collection of i.i.d. continuous random variables has their rank vector following a uniform distribution. Since the function a from (2) depends on its arguments only through their ranks, we have that, for $x_1, x_2 \in (0, 1)$,

$$g_X(x_1, x_2) = E[a(x_1, x_2, X_3, X_4)] = E[a(x_1, x_2, Y_3, Y_4)] = g_Y(x_1, x_2).$$

Applying Lemma 4.1, we have that

$$h_2((x_1, x_2), (y_1, y_2)) = \frac{2}{3}\ g_X(x_1, x_2)\ g_X(y_1, y_2)$$

and the proof is complete once the following claim is established:

$$g_X(x_1, x_2) = -3c(x_1, x_2), \quad x_1, x_2 \in (0, 1). \tag{20}$$

Letting $x_{(1)} = x_1 \wedge x_2 = \min\{x_1, x_2\}$ and $x_{(2)} = x_1 \vee x_2 = \max\{x_1, x_2\}$, we have

$$\begin{aligned} g_X(x_1, x_2) &= P(x_1, X_3 < x_2, X_4) + P(x_1, X_3 > x_2, X_4) \\ &\quad - P(x_1, x_2 < X_3, X_4) - P(x_1, x_2 > X_3, X_4) \\ &= P(x_{(1)}, X_3 < x_{(2)}, X_4) - P(x_{(2)} < X_3, X_4) - P(x_{(1)} > X_3, X_4) \\ &= P(x_{(1)}, X_3 < x_{(2)}, X_4) - (1 - x_{(2)})^2 - x_{(1)}^2. \end{aligned}$$

Moreover,

$$\begin{aligned} P(x_{(1)}, X_3 < x_{(2)}, X_4) &= P(x_{(1)} < X_4 \text{ and } X_3 < x_{(1)}) \\ &\quad + P(X_3 < X_4 \text{ and } x_{(1)} < X_3 < x_{(2)}) \end{aligned}$$

$$= x_{(1)}(1 - x_{(1)}) + \int_{x_{(1)}}^{x_{(2)}} P(x < X_4 \mid X_3 = x)\, \mathrm{d}x$$
$$= x_{(1)}(1 - x_{(1)}) + \int_{x_{(1)}}^{x_{(2)}} (1 - x)\, \mathrm{d}x$$
$$= x_{(1)}(1 - x_{(1)}) + x_{(2)} \left(1 - \frac{1}{2} x_{(2)}\right) - x_{(1)} \left(1 - \frac{1}{2} x_{(1)}\right).$$

We obtain that

$$g_X(x_1, x_2) = -1 - \frac{3}{2} x_{(1)}^2 - \frac{3}{2} x_{(2)}^2 + 3x_{(2)} = -1 - \frac{3}{2} x_1^2 - \frac{3}{2} x_2^2 + 3x_{(2)},$$

which is the claim from (20). □

Proof of Theorem 4.4. Let c be the kernel function for the Cramér-von Mises statistic, as defined in Lemma 4.3. The operator A_c is known to have eigenvalues $\frac{1}{j^2\pi^2}$ with corresponding eigenfunctions $\sqrt{2}\cos(\pi j x)$ for $j = 1, 2, \dots$ [13, Example 12.13]. Since $h_2((x_1, y_1), (x_2, y_2)) = 6c(x_1, x_2)c(y_1, y_2)$ by Lemma 4.3, it follows from Lemma 4.2 that $\frac{1}{6} h_2$ has eigenvalues $\{\frac{1}{\pi^4}\frac{1}{j^2 i^2} : (i, j) \in \mathbb{N}_+^2\}$ corresponding to orthonormal eigenfunctions $\{2\cos(\pi j x)\cos(\pi j y) : (i, j) \in \mathbb{N}_+^2\}$. The eigenvalues of h_2 are a multiple of 6 larger, with the same orthonormal eigenfunctions. We obtain from Theorem 2.3 that

$$nt^* \xrightarrow{d} \binom{4}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{\pi^4} \frac{6}{j^2 i^2} (\chi_{1,ij}^2 - 1) = \frac{36}{\pi^4} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{j^2 i^2} (\chi_{1,ij}^2 - 1)$$

where $\{\chi_{1,ij}^2 : i, j \in \mathbb{N}_+\}$ is a collection of i.i.d. χ_1^2 random variables. □

Proof of Theorem 4.5. Recall from Lemma 4.1 that we have the factorization

$$h_2((x_1, y_1), (x_2, y_2)) = \frac{2}{3}\, g_X(x_1, x_2)\, g_Y(y_1, y_2).$$

As in the proof of Theorem 4.4, finding the eigenvalues of h_2 requires only finding the eigenvalues of the operators A_{g_X} and A_{g_Y}. To obtain positive eigenvalues it will be useful to instead find the eigenvalues of A_{-g_X} and A_{-g_Y} which are simply the negation of the eigenvalues of A_{g_X} and A_{g_Y}. For notational simplicity let $k_X = -g_X$ and $k_Y = -g_Y$. Obtaining the eigenvalues of A_{k_X} and A_{k_Y} are two analogous problems and we thus discuss only A_{k_X}. We denote the support of X by $\{u_1, \dots, u_r\}$.

Let F and p be the cumulative density function and probability mass function corresponding to X respectively. Combining the first two probabilities in (8) and using that X_3 and X_4 are i.i.d. copies of X, the function $k_X(x_1, x_2) = -E[a(x_1, x_2, X_3, X_4)]$ can be written as

$$k_X(x_1, x_2)$$
$$= P(x_1 \wedge x_2, X_3 < x_1 \vee x_2, X_4) - P(x_1 \vee x_2 < X)^2 - P(x_1 \wedge x_2 > X)^2$$

$$= \left[(F(x_1 \wedge x_2) - p(x_1 \wedge x_2))^2 + (1 - F(x_1 \vee x_2))^2 \right]$$
$$- I(x_1 \neq x_2) \left[F(x_1 \wedge x_2)(1 - F(x_1 \wedge x_2)) + \sum_{x_1 \wedge x_2 < u_\ell < x_1 \vee x_2} p(u_\ell)(1 - F(u_\ell)) \right].$$

Finding the eigenvalues of A_{k_X} requires finding $\lambda \in \mathbb{R}$ and a function φ such that

$$\lambda \varphi(x) = E[k_X(x, X_2)\varphi(X_2)] = \sum_{j=1}^{r} p_X(u_j)\varphi(u_j)k_X(x, u_j), \tag{21}$$

for x in the support of X. Since the support is finite, (21) is a system of r linear equations in r unknowns $\varphi(u_1), \dots, \varphi(u_r)$. We recognize that the eigenvalues of A_{k_X} are the eigenvalues of the $r \times r$ matrix $\widetilde{R}^X$ whose (i, j)-th entry is $k_X(u_i, u_j)p_X(u_j)$.

Let K^X be the symmetric $r \times r$ matrix with (i, j)-th entry $k_X(u_i, u_j)$, and let $\text{diag}(p_X)$ be the diagonal $r \times r$ matrix whose diagonal entries are $p_X(u_1), \dots, p_X(u_r)$. Then $\widetilde{R}^X = K^X \, \text{diag}(p_X)$. Noting that $\widetilde{R}^X$ has same eigenvalues as the symmetric matrix $R^X = \text{diag}(p_X)^{1/2} K^X \, \text{diag}(p_X)^{1/2}$, we obtain that the eigenvalues of A_{k_X} are the eigenvalues of R^X, which is the matrix given by (9). Since the analogous fact holds for k_Y, an application of Lemma 4.2 and Theorem 2.3 completes the proof. □

References

[1] W. Bergsma and A. Dassios. A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20(2):1006–1028, 2014. MR3178526

[2] J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32:485–498, 1961. MR0125690

[3] S. S. Dhar, A. Dassios, and W. Bergsma. A study of the power and robustness of a new test for independence against contiguous alternatives. *Electron. J. Statist.*, 10(1):330–351, 2016. MR3466185

[4] E. B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, 11(3):739–745, 1983. MR0707925

[5] Y. Heller and R. Heller. Computing the Bergsma Dassios sign-covariance. Available at: arXiv:1605.08732, 2016.

[6] W. Hoeffding. A non-parametric test of independence. *Ann. Math. Statist.*, 19:546–557, 1948. MR0029139

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

[8] R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.

[9] M. L. Rizzo and G. J. Szekely. *energy: E-statistics (energy statistics)*, 2014. URL http://CRAN.R-project.org/package=energy. R package version 1.6.2.

[10] R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York, 1980. Wiley Series in Probability and Mathematical Statistics. MR0595165

[11] C. Spearman. The proof and measurement of association between two things. *Am. J. Psychol.*, 15:72–101, 1904.

[12] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35(6):2769–2794, 2007. MR2382665

[13] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. MR1652247

[14] L. Weihs. *TauStar: Efficient Computation and Testing of the Bergsma-Dassios Sign Covariance*, 2016. URL https://CRAN.R-project.org/package=TauStar. R package version 1.1.0.

[15] L. Weihs, M. Drton, and D. Leung. Efficient computation of the Bergsma–Dassios sign covariance. *Comput. Statist.*, 31(1):315–328, 2016.