Subhra Sankar Dhar, Angelos Dassios and Wicher Bergsma

A study of the power and robustness of a new test for independence against contiguous alternatives

Article (Published version) (Refereed)

Original citation:
Dhar, Subhra Sankar, Dassios, Angelos and Bergsma, Wicher (2016) *A study of the power and robustness of a new test for independence against contiguous alternatives*. Electronic Journal of Statistics, 10 (1). pp. 330-351. ISSN 1935-7524
DOI: 10.1214/16-EJS1107

Electronic Journal of Statistics
Vol. 10 (2016) 330–351
ISSN: 1935-7524
DOI: 10.1214/16-EJS1107

A study of the power and robustness of a new test for independence against contiguous alternatives

Subhra Sankar Dhar

IIT Kanpur, India
e-mail: subhra@iitk.ac.in

Angelos Dassios and Wicher Bergsma

London School of Economics and Political Science, UK
e-mail: A.Dassios@lse.ac.uk; W.P.Bergsma@lse.ac.uk

Abstract: Various association measures have been proposed in the literature that equal zero when the associated random variables are independent. However many measures, (e.g., Kendall's tau), may equal zero even in the presence of an association between the random variables. In order to overcome this drawback, Bergsma and Dassios (2014) proposed a modification of Kendall's tau, (denoted as τ^*), which is non-negative and zero *if and only if* independence holds. In this article, we investigate the robustness properties and the asymptotic distributions of τ^* and some other well-known measures of association under null and contiguous alternatives. Based on these asymptotic distributions under contiguous alternatives, we study the asymptotic power of the test based on τ^* under contiguous alternatives and compare its performance with the performance of other well-known tests available in the literature.

MSC 2010 subject classifications: Primary 62G35, 62G10; secondary 62G20.
Keywords and phrases: Contiguous alternatives, Distance covariance, Kendall's tau, Pitman efficacy, Robustness properties, Test for independence.

Received June 2015.

1. Introduction

Since the early part of the last century, several measures of association have been proposed to detect the association between random variables. Some of the most popular are Kendall's τ (see Kendall (1938), Spearman's ρ (see Spearman (1904), Hoeffding's coefficient (see Hoeffding (1948), Blum-Kiefer-Rosenblatt's coefficient (see Blum, Kiefer, and Rosenblatt (1961), distance covariance (see Székely, Rizzo, and Bakirov (2007) and Kolmogorov—Smirnov and Cramer von Mises tests (e.g., see Serfling (1980). Among these tests, Kendall's τ and Spearman's ρ have effective representations in terms of the sign function. For two random variables X and Y, Kendall's τ is defined as E sign$\{(X_1 - X_2)(Y_1 - Y_2)\}$, and Spearman's ρ is defined as E sign$\{(X_1 - X_2)(Y_1 - Y_3)\}$, where (X_1, Y_1),

(X_2, Y_2) and (X_3, Y_3) are independent replications of (X, Y) (see, e.g., Gibbons and Chakraborti (2011), and $\text{sign}(x) = x/|x|$, if $x \neq 0$, and $\text{sign}(x) = 0$, if $x = 0$. It follows from the definitions of τ and ρ that both τ, $\rho \in [-1, 1]$, and $\tau = \rho = 0$ if $X \perp\!\!\!\perp Y$. For given data $(x_1, y_1), \ldots, (x_n, y_n)$, the sample versions of τ and ρ can be defined as $\tau_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}\{(x_i - x_j)(y_i - y_j)\}$ and $\rho_n = \frac{1}{\binom{n}{3}} \sum_{1 \leq i < j < k \leq n} \text{sign}\{(x_i - x_j)(y_i - y_k)\}$, respectively.

However, *τ and ρ may equal zero even in the presence of association between X and Y*. In order to make this relationship equivalent (i.e., the measure of association $= 0$ *if and only if* $X \perp\!\!\!\perp Y$), Bergsma and Dassios (2014) proposed a new measure (denoted as τ^*), which is defined as $\tau^* = Ea(X_1, X_2, X_3, X_4)a(Y_1, Y_2, Y_3, Y_4)$, where $a(z_1, z_2, z_3, z_4) = \text{sign}(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$, and (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) and (X_4, Y_4) are independent replications of (X, Y). Bergsma and Dassios (2014) showed that $\tau^* \geq 0$, and $\tau^* = 0 \Leftrightarrow X \perp\!\!\!\perp Y$. It also follows from the definition of Kendall's τ that $\tau^2 = Es(X_1, X_2, X_3, X_4)s(Y_1, Y_2, Y_3, Y_4)$, where $s(z_1, z_2, z_3, z_4) = \text{sign}(|z_1 - z_2|^2 + |z_3 - z_4|^2 - |z_1 - z_3|^2 - |z_2 - z_4|^2)$. Note that the form of τ^2 is similar to the form of τ^* though $\tau^2 = 0$ does not satisfy the *if and only if* condition as we mentioned earlier. For the given data $(x_1, y_1), \ldots, (x_n, y_n)$, the sample version of τ^* (denoted as τ_n^*) can be defined as $\tau_n^* = \frac{1}{\binom{n}{4}} \sum_{1 \leq i < j < k < l \leq n} a(x_i, x_j, x_k, x_l)a(y_i, y_j, y_k, y_l)$.

Another recently popularized measure of association is distance covariance (see Székely et al. (2007), which is defined as $\text{dcov} = E||X_1 - X_2||\ ||Y_1 - Y_2|| + E||X_1 - X_2||\ E||Y_1 - Y_2|| - 2E||X_1 - X_2||\ ||Y_1 - Y_3||$, where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are independent replications of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^q$, $p, q \geq 1$. It is straightforward to show that $\text{dcov} = \frac{1}{4} Eh(X_1, X_2, X_3, X_4)h(Y_1, Y_2, Y_3, Y_4)$, where $h(z_1, z_2, z_3, z_4) = |z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|$, and (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4) are independent replications of (X, Y). In addition, one can also show that $\text{dcov} = \frac{1}{c_p c_q} \int_{\mathbb{R}^p \times \mathbb{R}^q} \frac{|\psi_{X,Y}(s,t) - \psi_X(s)\psi_Y(t)|^2}{||t||^{1+p}||s||^{1+q}} ds dt$ (see Székely et al. (2007)), where $\psi_{X,Y}$, ψ_X and ψ_Y are the characteristic functions of (X, Y), X and Y respectively, and this definition implies that $\text{dcov} = 0 \Leftrightarrow X \perp\!\!\!\perp Y$. For the given data $(x_1, y_1), \ldots, (x_n, y_n)$, the sample version of dcov is defined as $\text{dcov}_n = \frac{1}{\binom{n}{4}} \sum_{1 \leq i < j < k < l \leq n} \frac{1}{4} h(x_i, x_j, x_k, x_l)h(y_i, y_j, y_k, y_l)$. However, it is expected that dcov is not a robust measure of association since it is moment based, whereas τ and τ^* are expected to be robust against the outliers since these measures are based on the ranks or the positions of the observations. Due to this reason, it is also expected that the test based on τ_n^* will be more powerful than the test based on dcov_n when the null and the alternative distributions are associated with two *well separated distinct* populations, i.e., the data from one population can be considered as the outliers relative to the data cloud formed by the observations obtained from the other population.

Along with the issue of robustness of different measures of independence, it is also of interest whether we can determine if the pair of random variables are independent or not based on τ, τ^* and dcov. In order to investigate this testing of hypothesis problem, one should ideally carry out the tests based on the exact distributions of τ_n, τ_n^* and dcov_n. However, since the exact distributions of τ_n, τ_n^* and dcov_n are not tractable, we estimate the size and the power of the tests based on the asymptotic distributions of τ_n, τ_n^* and dcov_n. In addition, since both the tests based on τ_n^* and dcov_n are consistent tests against fixed alternative, here we investigate the asymptotic powers of the tests under contiguous alternatives. In short, under the condition of *contiguity*, a limit law Q_n of random vectors $X_n : \Omega_n \to \mathbb{R}^k$, $k \geq 1$ can be obtained from a suitable other limit law P_n, where Q_n and P_n are probability measures defined on $(\Omega_n, \mathcal{A}_n)$. The more technical issues related to contiguity will be discussed at the beginning of Section 3.

The rest of the article is organized as follows. In Section 2, we study the robustness of the aforementioned measures of association. In Section 3, we obtain the asymptotic distributions of τ_n, τ_n^* and dcov_n under null and contiguous alternatives, and based on these results, we investigate the asymptotic powers of the tests based on these statistics. Section 4 contains some concluding remarks. All technical details appear in the appendix.

2. Robustness study

(Huber, 2011, p. 9, 11) discusses a concept of maximum bias to investigate the robustness of the estimator (or the corresponding functional), which is based on a contamination neighborhood. The maximum bias of $T(\cdot)$ is defined as $b_1(\epsilon) = \sup_{F \in \mathcal{P}_\epsilon} |T(F) - T(F_0)|$, for any small $\epsilon > 0$, where $\mathcal{P}_\epsilon = \{F : F = (1-\epsilon)F_0 + \epsilon H, H \in \mathcal{M}\}$, F_0 is the true distribution function, and $\mathcal{M}$ is the collection of probability measures such that the map $F \to \int \psi dF$ from $\mathcal{M}$ into $\mathbb{R}$ is continuous whenever ψ is bounded and continuous. Motivated by the concept of maximum bias, we define a new measure $b(\beta; T(F_0))$ as follows. Let H be the dirac measure, i.e., $H_{X,Y} = \delta_{X,Y}(h,k)$, where $\delta_{X,Y}(h,k) = 1$, if $(X,Y) = (h,k)$, and $\delta_{X,Y}(h,k) = 0$, if $(X,Y) \neq (h,k)$, and let F_0 be the joint distribution function of (X,Y), whose associated random vector has independent components. Finally, $b(\beta; T(F_0))$ is defined as

$$b(\beta; T(F_0)) = \lim_{h,k\to\infty} |T((1-\beta)F_0 + \beta\delta_{X,Y}(h,k)) - T(F_0)|.$$

In other words, $b(\beta; T(F_0))$ measures the effect on $T(F_0)$ of an arbitrary large observation with mass β.

Remark 1. *It is also appropriate to mention here that one can define* $b(\beta; T(F_0))$ *when* $h, k \to \pm\infty$*, and in view of the fact that* τ*,* τ^* *and* dcov *are based on the absolute values of the differences between the observations, the values of* $b(\beta; .)$ *measure for* $\tau(F_0)$*,* $\tau^*(F_0)$ *and* $\mathrm{dcov}(F_0)$ *will remain the same when* $h, k \to -\infty$*. However, for the sake of simplicity, we assume* $h, k \to \infty$ *throughout the paper unless mentioned otherwise.*

The following theorem states the behaviour of $b(\beta; \tau^*(F_0))$.

Theorem 1. *Let* F_0 *be a joint distribution function of* (X, Y)*, whose associated marginal distribution functions are* G_X *and* H_Y *of* X *and* Y*, respectively and in addition,* $F_0 = G_X H_Y$*. Then, for any* $\beta < 1/2$*, we have* $b(\beta; \tau^*(F_0)) = 4\beta^2(1-$

$\beta)^2$, and consequently, $b(\beta;\tau^*(F_0)) < 1/4$ for any $\beta < 1/2$. Here $\tau^*(F_0) = E_{F_0}\{a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)\}$, where (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) and (X_4,Y_4) are independent replications of (X,Y), and for any z_1, z_2, z_3 and z_4, $a(z_1,z_2,z_3,z_4) = \text{sign}(|z_1-z_2|+|z_3-z_4|-|z_1-z_3|-|z_2-z_4|)$.

Theorem 1 implies that in the presence of $\beta \in [0,1/2)$ proportion outliers in the data, the bias of the functional τ^* evaluated at F_0 will be bounded by β^2. In fact, strictly speaking, the bias will be bounded by $1/4$. In other words, the bias will not *break down* to 1 even in the presence of arbitrarily large outliers. Also, in view of the fact that $\tau_n^* \to \tau^*$ in probability as $n \to \infty$, the bias of τ_n^* will be bounded by $1/4$ in probability when the data is obtained from the joint distribution function having independent marginal distribution functions.

Proposition 1 discusses the behaviour of $b(\beta;\tau(F_0))$ and $b(\beta;\text{dcov}(F_0))$.

Proposition 1. *Under the conditions of Theorem 1, for any $\beta < 1/2$, we have $b(\beta;\tau(F_0)) = 4\beta^2(1-\beta)^2$, and consequently, $b(\beta;\tau(F_0)) < 1/4$ for any $\beta < 1/2$. Here $\tau(F_0) = E_{F_0}[\text{sign}\{(X_1-X_2)(Y_1-Y_3)\}]$, where (X_1,Y_1) and (X_2,Y_2) are independent replications of (X,Y). Under the same conditions, for any $\beta < 1/2$, we have $b(\beta;\text{dcov}(F_0)) = \infty$. Here $\text{dcov}(F_0) = \frac{1}{4}E_{F_0}\{h(X_1,X_2,X_3,X_4)h(Y_1,Y_2,Y_3,Y_4)\}$, where (X_1,Y_1), (X_2,Y_2), (X_3,Y_3), (X_4,Y_4) are independent replications of (X,Y), and for any z_1, z_2, z_3 and z_4, $h(z_1,z_2,z_3,z_4) = |z_1-z_2|+|z_3-z_4|-|z_1-z_3|-|z_2-z_4|$.*

The assertion in Proposition 1 implies that τ is also a robust measure in the sense of having bounded $b(\beta;\tau(F_0))$, whereas unlike $b(\beta;\tau(F_0))$ and $b(\beta;\tau^*(F_0))$, $b(\beta;\text{dcov}(F_0))$ is unbounded. This fact implies that distance covariance is *non-robust* against the outliers. As we have mentioned in the Introduction, the non-robustness of distance covariance is expected to be reflected in the asymptotic power study, which will be fully discussed in the forthcoming section.

3. Asymptotic power study under contiguous alternatives

Besides the issue of robustness, since the tests based on both τ_n^* and dcov_n are consistent (i.e., the power of the test tends to one as the sample size tends to infinite), a natural question is how the asymptotic powers of the tests based on τ_n^* and dcov_n compare with other well-known tests (e.g., a test based on τ_n) under contiguous alternatives (e.g., see Hajek, Sidak, and Sen (1999), p. 249). Precisely, the sequence of probability measures Q_n is contiguous with respect to the sequence of probability measures P_n if $P_n(A_n) \to 0$ implies that $Q_n(A_n) \to 0$ for every sequence of measurable sets A_n, where $(\Omega_n, \mathcal{A}_n)$ is the sequence of measurable spaces, and P_n and Q_n are two probability measures defined on $(\Omega_n, \mathcal{A}_n)$. In order to characterise the contiguity in terms of the asymptotic behaviour of the likelihood ratios between P_n and Q_n, Le Cam proposed some results popularly known as Le Cam's Lemma (e.g., see Hajek et al. (1999)). A consequence of Le Cam's first lemma is that the sequence Q_n will be contiguous with respect to the sequence P_n if $\log\frac{Q_n}{P_n}$ asymptotically follow a Gaussian distribution with mean $= -\frac{\sigma^2}{2}$ and Variance $= \sigma^2$ under P_n (e.g., see Hajek et al. (1999), p. 253, Corollary to Le Cam's first Lemma)), where $\sigma > 0$ is a constant, and we use this fact to establish contiguity in this article (see the proof of Theorem 2).

Suppose that we now want to test $H_0 : F_0 = G_X H_Y$, where F_0 is the joint distribution function of (X, Y) with the associated marginal distribution functions of X and Y being G_X and H_Y, respectively, and we consider a sequence of contiguous or local alternatives $H_n : F_n = (1-\gamma/\sqrt{n})F_0 + (\gamma/\sqrt{n})K$ for a fixed $\gamma > 0$ and $n = 1, \cdots,$. Here we should point out that A_n is a sequence of sets, which is changing over n along with its σ-field $\mathcal{A}_n$, and for that reason, it does *not follow directly* from the definition of contiguity that F_n is contiguous with respect to F_0. In Theorem 2, based on Le Cam's first lemma, we establish that the alternatives H_n will be contiguous alternatives under certain conditions.

In order to carry out the tests based on τ_n, τ_n^* and dcov_n, one needs to know the distributions (or an approximation of the distributions) of these estimators. In this context, note that τ_n, τ_n^* and dcov_n are U-statistics (e.g., see Lee (1990)) and to derive the asymptotic distributions of them, one needs to know the order of degeneracy of each τ_n, τ_n^* and dcov_n. For the sake of completeness, the definition of U-statistic and its order of degeneracy are given below. For the given data $\mathcal{X} = \{x_1, \ldots, x_n\}$, $U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \ldots < i_m} k(x_{i_1}, \ldots, x_{i_m})$ is said to be a U-statistic of order m with kernel $k(\cdot)$ having the order of degeneracy $= l$ if $Var(E_{X_{l+1},\ldots,X_m} k(X_1, \ldots, X_l, X_{l+1}, \ldots, X_m)) = 0$ but $Var(E_{X_{l+2},\ldots,X_m} k(X_1, \ldots, X_{l+1}, X_{l+2}, \ldots, X_m)) > 0$, i.e., in other words, $E_{X_{l+1},\ldots,X_m} k(x_1, \ldots, x_l, X_{l+1}, \ldots, X_m) = 0$, for all $x_1, \ldots, x_l$. The statistic τ_n has the order of degeneracy $= 0$, whereas that of τ_n^* and dcov_n are of order 1 (see the proofs of Theorems 2, 3 and 4). Here it should be further pointed out that $0 = \delta_0^2 \leq \delta_1^2 \leq \ldots \leq \delta_m^2$, where $\delta_l^2 = Var(E_{X_{l+1},\ldots,X_m} k(X_1, \ldots, X_l, X_{l+1}, \ldots, X_m))$ for $l = 1, \ldots, m$ (see, e.g., Serfling (1980), p. 182), which implies that for all $k \geq l$, $\delta_k^2 > 0$ when $\delta_l^2 > 0$. This fact ensures the uniqueness of the order of degeneracy of U-statistic in view of the definition of the order of degeneracy. The connection between the rate of convergence of U-statistic and its order of degeneracy will be discussed in Remark 3. In Theorems 2, 3 and 4, we describe the asymptotic behaviour of τ_n, τ_n^* and dcov_n respectively under contiguous alternatives H_n.

Theorem 2. *Assume that F_0 and K have Lebesgue densities f_0 and k, respectively, and $E_{f_0}\{\frac{k}{f_0} - 1\}^2 < \infty$. Then, the sequence of alternatives H_n is contiguous to H_0. Moreover, under H_n, $\sqrt{n}(\tau_n - \tau)$ converges weakly to a Gaussian distribution with mean μ_1 and variance σ_1^2, where*

$$\mu_1 = 2\gamma \int_{\infty}^{\infty}\int_{\infty}^{\infty} \left[2 \int_{x}^{\infty}\int_{y}^{\infty} f_0(u,v)dudv + 2 \int_{-\infty}^{x}\int_{-\infty}^{y} f_0(u,v)dudv - 1 \right] k(x,y)dxdy$$

and

$$\sigma_1^2 = 4 \int_{\infty}^{\infty}\int_{\infty}^{\infty} \left[2 \int_{x}^{\infty}\int_{y}^{\infty} f_0(u,v)dudv + 2 \int_{-\infty}^{x}\int_{-\infty}^{y} f_0(u,v)dudv - 1 \right]^2 f_0(x,y)dxdy.$$

Here note that $E_{f_0}(\log \frac{k}{f_0}) = E_{f_0} \log(1-(1-\frac{k}{f_0})) \approx -\frac{1}{2}E_{f_0}(1-\frac{k}{f_0})^2$ as $E_{f_0}(1-\frac{k}{f_0}) = 0$, and hence, $-\frac{1}{2}E_{f_0}(1-\frac{k}{f_0})^2$ is essentially the first order approximation of an entropy $E_{f_0}(\log \frac{k}{f_0})$ that measures dissimilarity between two densities f_0 and k. In other words, $E_{f_0}(1-\frac{k}{f_0})^2$ is the *mean square contingency* (see (Rényi, 1959, p. 446)) of f_0 and k. Further, we should mention that if $k = f_0$, we have $\frac{k}{f_0} - 1 = 0$, i.e., k and f_0 are similar. At the same time, larger values of $\frac{k}{f_0} - 1$ indicate that k and f_0 are more dissimilar. To summarize, Theorem 2 asserts that the sequence of alternatives H_n will be contiguous with respect to H_0 when the *mean square contingency* of f_0 and k is finite.

To prove Theorem 2, Le Cam's third lemma is used to obtain the asymptotic normality of $\sqrt{n}\{\tau_n^* - \tau\}$ under H_n, and Le Cam's third lemma uses the fact that $\log L_n$ converges weakly to a random variable associated with a normal distribution having certain location and scale parameters (see the proof of Theorems 2). We should point out that the asymptotic normality of $\log L_n$ is a sufficient condition but not a necessary condition to establish the contiguity of Q_n with respect to P_n. Instead of Le Cam's third lemma, one can also follow Behnen and Neuhaus (1975)'s approach based on a specific truncation method for contiguity of the density functions associated with H_n with respect to the density function associated with H_0. Also, Behnen (1971) investigated the asymptotic relative efficiency of some tests for independence against general contiguous alternatives of positive quadrant dependence. However, neither Behnen (1971) nor Behnen and Neuhaus (1975) considered the distribution functions associated with H_n as a mixture distribution, as we consider here. Recently, Banerjee (2005) studied the behaviour of the likelihood ratio statistics for testing a finite dimensional parameter under local contiguous hypotheses. To obtain the local (or contiguous) alternatives, he perturbed the null hypothesized parameter, which is different from the perturbance on the distribution function considered by us.

Note that the sequence of contiguous alternatives H_n coincide with the null hypothesis H_0 when $\gamma = 0$, and hence, the asymptotic distribution of $\sqrt{n}(\tau_n - \tau)$ under H_0 directly follows from the assertion in Theorem 2 by choosing $\gamma = 0$. Corollary 1 states the asymptotic distribution of $\sqrt{n}(\tau_n - \tau)$ under H_0.

Corollary 1. *Assume that F_0 has density function f_0. Then, under H_0, $\sqrt{n}(\tau_n - \tau)$ converges weakly to a Gaussian distribution with mean zero and variance σ_1^2, where σ_1^2 is the same as defined in Theorem 2.*

Theorem 3. *Assume the same conditions on F_0 and K as mentioned in Theorem 2. Then, under H_n, $n(\tau_n^* - \tau^*)$ converges weakly to $\sum_{i=1}^{\infty} \lambda_i\{(Z_i + a_i)^2 - 1\}$, where Z_i's are i.i.d. $N(0,1)$ random variables, and λ_i's are the eigenvalues associated with $l(x,y) = E\{\text{sign}(|X_1 - X_2| + |X_3 - X_4| - |X_1 - X_3| - |X_2 - X_4|) \times \text{sign}(|Y_1 - Y_2| + |Y_3 - Y_4| - |Y_1 - Y_3| - |Y_2 - Y_4|) | X_1 = x, Y_1 = y\}$. Here (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) and (X_4, Y_4) are i.i.d. bivariate random vectors, and*

$$a_i = \gamma \int \left\{ \frac{k(x,y)}{f_0(x,y)} - 1 \right\} g_i(x) g_i(y) f_{X,Y}\, dxdy,$$

where $g_i(x)$ and $g_i(y)$ are the eigenfunctions such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(x,y)\prod_{i=2}^{4} g_k(X_i)g_k(Y_i)d(\prod_{i=2}^{4} F_{0;X_i,Y_i}) = \lambda_k g_k(x)g_k(y) \text{ for all } (x,y).$$

Theorem 4. *Assume the same conditions on F_0 and K as mentioned in Theorem 2. Then, under H_n, $n(\mathrm{dcov}_n - \mathrm{dcov})$ converges weakly to $\sum_{i=1}^{\infty}\lambda_i^*\{(Z_i^* + a_i^*)^2 - 1\}$, where Z_i^*'s are i.i.d. $N(0,1)$ random variables, and λ_i^*'s are the eigenvalues associated with $l^*(x,y) = E\{(|X_1 - X_2| + |X_3 - X_4| - |X_1 - X_3| - |X_2 - X_4|) \times (|Y_1 - Y_2| + |Y_3 - Y_4| - |Y_1 - Y_3| - |Y_2 - Y_4|)|X_1 = x, Y_1 = y\}$. Here (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) and (X_4,Y_4) are i.i.d. bivariate random vectors, and*

$$a_i = \gamma\int\left\{\frac{k(x,y)}{f_0(x,y)} - 1\right\} g_i^*(x)g_i^*(y)f_{X,Y}\,dxdy,$$

where $g_i^*(x)$ and $g_i^*(y)$ are the eigenfunctions such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l^*(x,y)\prod_{i=2}^{4} g_i^*(X_i)g_i^*(Y_i)d(\prod_{i=2}^{4} F_{0;X_i,Y_i}) = \lambda_k^* g_k^*(x)g_k^*(y) \text{ for all } (x,y).$$

Here again, when $\gamma = 0$, the sequence of contiguous alternatives H_n for $n = 1, \cdots$ coincide with the null hypothesis H_0, and as a consequence, one can derive the asymptotic distributions of $n(\tau_n^* - \tau^*)$ and $n(\mathrm{dcov}_n - \mathrm{dcov})$ under H_0 from their asymptotic distributions under H_n when $\gamma = 0$. Corollaries 2 and 3 state the asymptotic distributions of $n(\tau_n^* - \tau^*)$ and $n(\mathrm{dcov}_n - \mathrm{dcov})$ under H_0, respectively.

Corollary 2. *Assume the same conditions on F_0 as mentioned in Theorem 2. Then, under H_0, $n(\tau_n^* - \tau^*)$ converges weakly to $\sum_{i=1}^{\infty}\lambda_i(Z_i^2 - 1)$, where Z_i's are i.i.d. $N(0,1)$ random variables, and λ_i's are the eigenvalues associated with $l(x,y) = E\{\mathrm{sign}(|X_1 - X_2| + |X_3 - X_4| - |X_1 - X_3| - |X_2 - X_4|) \times \mathrm{sign}(|Y_1 - Y_2| + |Y_3 - Y_4| - |Y_1 - Y_3| - |Y_2 - Y_4|)|X_1 = x, Y_1 = y\}$. Here (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) and (X_4,Y_4) are i.i.d. bivariate random vectors, and $g_i(x)$ and $g_i(y)$ are the eigenfunctions such that*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(x,y)\prod_{i=2}^{4} g_k(X_i)g_k(Y_i)d(\prod_{i=2}^{4} F_{0;X_i,Y_i}) = \lambda_k g_k(x)g_k(y) \text{ for all } (x,y).$$

Corollary 3. *Assume the same conditions on F_0 as mentioned in Theorem 2. Then, under H_0, $n(\mathrm{dcov}_n - \mathrm{dcov})$ converges weakly to $\sum_{i=1}^{\infty}\lambda_i^*\{Z_i^{*2} - 1\}$, where Z_i^*'s are i.i.d. $N(0,1)$ random variables, and λ_i^*'s are the eigenvalues associated with $l^*(x,y) = E\{(|X_1 - X_2| + |X_3 - X_4| - |X_1 - X_3| - |X_2 - X_4|) \times (|Y_1 - Y_2| + |Y_3 - Y_4| - |Y_1 - Y_3| - |Y_2 - Y_4|)|X_1 = x, Y_1 = y\}$. Here (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) and (X_4,Y_4) are i.i.d. bivariate random vectors, and $g_i^*(x)$ and $g_i^*(y)$*

are the eigenfunctions such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l^*(x,y)\prod_{i=2}^{4} g_k^*(X_i)g_k^*(Y_i)d(\prod_{i=2}^{4} F_{0;X_i,Y_i}) = \lambda_k^* g_k^*(x)g_k^*(y) \ for\ all\ (x,y).$$

The assertion in Corollary 1 implies that $\sqrt{n}(\tau_n - \tau) = O_p(1)$, which follows from Prohorov's theorem (e.g., see (Van der Vaart, 2000, p. 8), and consequently, we have $\tau_n - \tau = o_p(1)$, which ensures that τ_n is a consistent estimator of τ. Similarly, along with a straightforward application of Prohorov's theorem (e.g., see (Van der Vaart, 2000, p. 8), it follows from Corollaries 2 and 3 that $n(\tau_n^* - \tau^*) = O_p(1)$ and $n(\text{dcov}_n - \text{dcov}) = O_p(1)$, respectively. These facts imply that τ_n^* and dcov_n are consistent estimators of τ^* and dcov respectively.

Remark 2. *It is appropriate to mention here that one can directly establish the results related to the consistency of τ_n, τ_n^* and dcov_n using the results on the consistency of U-statistics. Among these three estimators τ_n, τ_n^* and dcov_n, τ_n is a non-degenerate U-statistic, whereas τ_n^* and dcov_n are degenerate U-statistics of order $= 1$. The exact variance expressions of non-degenerate and degenerate U-statistics are given in p. 183 (Lemma A), and in p. 189 in Serfling (1980) respectively, and those variance terms converge to zero as $n \to \infty$ (see p. 183 in (iii) of Lemma A in Serfling (1980)). These facts establish the consistency of τ_n, τ_n^* and dcov_n to its population counterpart.*

Remark 3. *The rates of convergence of τ_n, τ_n^* and dcov_n also follow from the results related to the rate of convergence of the U-statistic. Based on the well-known projection method of the U-statistic (see, e.g., Section 5.3.4 in Serfling (1980), pp. 189–190), one can derive directly the rate of convergence of τ_n when $c = 1$ in the expression given in 5.3.4 in Serfling (1980) and that of τ_n^* and dcov_n when $c = 2$ in that expression. The aforementioned choices of c depend on the order of degeneracy of the corresponding U-statistic. In other words, this fact gives us an idea on how the rate of convergence of U-statistic is associated with its order of degeneracy. To summarize, for a U-statistic with the order of degeneracy $= p$, the rate of convergence will be of $n^{\frac{p+1}{2}}$, where n is the sample size, and p is an integer (see, e.g., Section 5.3.4 in Serfling (1980), pp. 189–190).*

Remark 4. *We would like to end this section with a discussion of the eigenvalues and the eigenfunctions, which are associated with the asymptotic distributions of τ_n^* and dcov_n stated in Theorems 3 and 4. In view of the non-zero order of degeneracy of τ_n^* and dcov_n, the eigenvalues and the eigenfunctions are involved in the asymptotic distributions of them (see, e.g., Section 5.5.2 in Serfling (1980), pp. 193–194). Further, using a spectral decomposition of the kernels $l(x,y)$ and $l^*(x,y)$, we have $l(x,y) = \sum_{k=1}^{\infty} \lambda_k g_k(x)g_k(y)$ and $l^*(x,y) = \sum_{k=1}^{\infty} \lambda_k^* g_k^*(x)g_k^*(y)$, which hold true in the L_2-sense. Here, $g_k(\cdot)$s are orthonormal eigenfunctions and λ_ks are the corresponding eigenvalues of the integral equation on $l(x,y)$ described in the statement of Theorem 3. Similarly, $g_k^*(\cdot)$s are orthonormal eigenfunctions and λ_k^*s are the corresponding eigen-*

values of the integral equation on $l^*(x,y)$ *described in the statement of Theorem 4. In addition, the orthonormality of* $g_k(\cdot)$ *implies that* $E\{g_k^2(X)\} = 1$ *and* $E\{g_k(X)g_{k'}(X)\} = 0$ *for all* $k \neq k'$*. Similarly, due to the same reason, we have* $E\{g_k^{*2}(X)\} = 1$ *and* $E\{g_k^*(X)g_{k'}^*(X)\} = 0$ *for all* $k \neq k'$*. Moreover, it is here appropriate to mention that for* $n(\mathrm{dcov}_n - \mathrm{dcov})$*, Bergsma (2006) listed the exact forms of the eigenvalues and the eigenfunctions for various distributions, and as a result, any asymptotic inference based upon Theorem 4 and Corollary 3 is feasible. However, the exact forms of the eigenvalues and the eigenfunctions associated with the asymptotic distributions of* $n(\tau_n^* - \tau^*)$ *are not yet available in the literature.*

3.1. Computation: Implementation of the tests and some examples

Theorem 2 helps us to compute the asymptotic power of the test based on τ_n for different values of γ, and the asymptotic critical value at $\alpha\%$ level of significance (denote it as $c_1(\alpha)$) can be obtained from the $(1-\alpha)\%$ quantile of the Gaussian distribution described in Corollary 1. Similarly, Theorems 3 and 4 enable us to compute the asymptotic power of the tests based on τ_n^* and dcov_n, and the corresponding asymptotic critical values (denoted as $c_2(\alpha)$ and $c_3(\alpha)$ respectively) can be obtained from the $(1-\alpha)\%$ quantile of the distributions described in Corollary 2 and Corollary 3 respectively. However, since the infinite sum of the weighted chi-squared distribution (see Theorems 3 and 4 and Corollary 2 and 3) with weights as the eigenvalues of the kernels associated with τ^* (or dcov), is not easily tractable in practice, it becomes difficult to have quantiles of this distribution. In order to overcome this problem related to infinitely many eigenvalues and the infinite sum, we approximate the kernel function at $n_1 \times n_1$ many marginal quantile points and compute the eigenvalues of $n_1 \times n_1$ finite-dimensional matrix associated with the kernel function. The (i,j)-th element of the matrix is the $(i/n_1, j/n_1)$-th marginal quantile (see Babu and Rao (1988)) of the joint distribution associated with the bivariate random vector (X,Y), where $i = 1,\ldots,n_1$ and $j = 1,\ldots,n_1$. Then, we generate a large sample with size n_2 from that approximated finite sum of the weighted chi-squared distribution, and the $(1-\alpha)\%$-th quantile of that sample is taken as the approximated value of the asymptotic critical value at $\alpha\%$ level of significance. Similarly, in order to compute power, we approximate the infinite sum of the weighted chi-squared distributions, described in Theorems 3 and 4, by an appropriate finite sum of the chi-squared distributions. We simulate a large sample with size n_3 from the approximated distributions, and finally, the proportion of the observations in the sample larger than the approximated critical value, is considered to be the value of the asymptotic power. Also, for distance covariance, we carry out an alternative procedure based on the exact forms of the first four eigenvalues, which essentially explains more than 90% variation (see Bergsma (2006)) and the corresponding eigenfunctions. The results obtained by this procedure are nearly the same as the reported results. In the asymptotic power studies of different tests, we consider $n_1 = 10$, $n_2 = 100$ and $n_3 = 100$ unless mentioned otherwise.

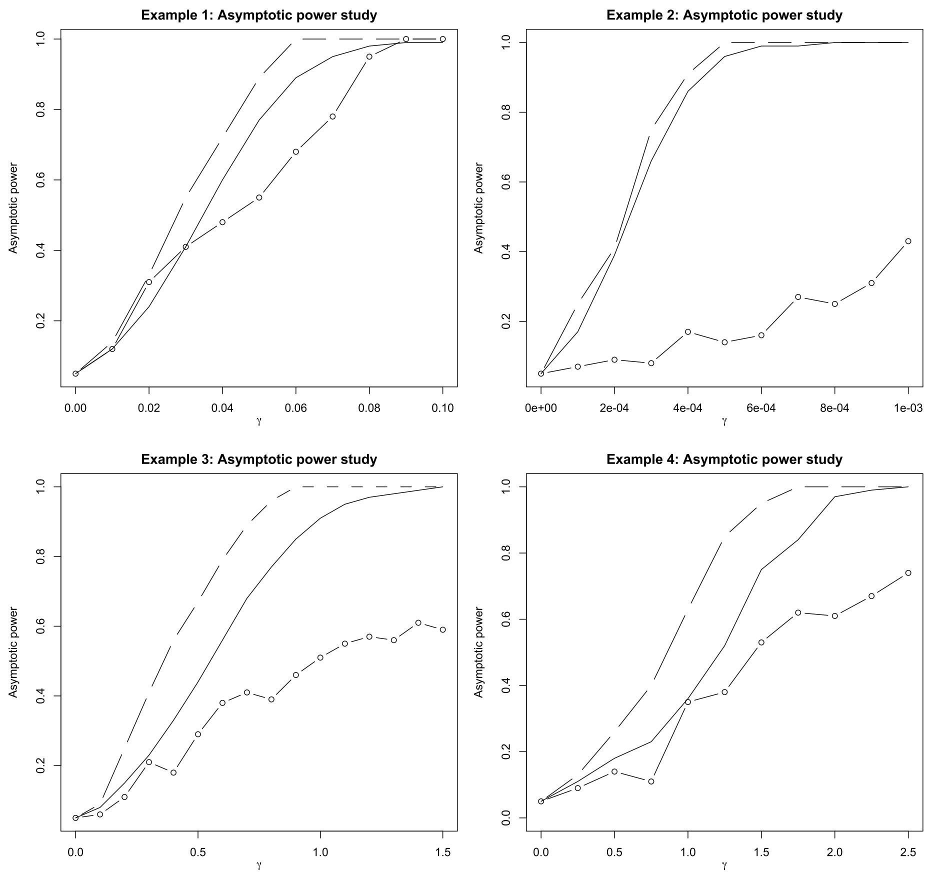

Fig 1. *The asymptotic power of the test based on* τ_n *(solid curve* $-$*), the test based on* τ_n^* *(lined curve* $-\ \ -$*) and the test based on* dcov_n *(dotted line curve* $-o-$*) for different values of* γ.

In the following examples, we compute the asymptotic power of the tests based on τ_n, τ_n^* and dcov_n for different values of γ with various choices of f_0 and k. All results are summarized in Figure 1.

Example 1. Consider

$$\begin{aligned} f_0(x,y) &= 1 \qquad \text{if} \qquad (x,y) \in [0,1]^2 \\ &= 0 \qquad \text{if} \qquad (x,y) \notin [0,1]^2, \end{aligned}$$

and

$$\begin{aligned} k(x,y) &= 1 \qquad \text{if} \qquad (x,y) \in [2,3]^2 \\ &= 0 \qquad \text{if} \qquad (x,y) \notin [2,3]^2. \end{aligned}$$

The results are reported in Table 1.

Table 1

The results for Example 1: The asymptotic power of the different tests for different values of γ. For different values of γ, the value within each cell of the second, the third, the fourth, the sixth, the seventh and the eighth rows denote the asymptotic power of the corresponding test at 5% level of significance under contiguous alternatives.

γ	0	0.01	0.02	0.03	0.04	0.05
Test based on τ_n	0.05	0.12	0.24	0.41	0.60	0.77
Test based on τ_n^*	0.05	0.14	0.33	0.55	0.72	0.89
Test based on dcov_n	0.05	0.12	0.31	0.41	0.48	0.55
γ	0.06	0.07	0.08	0.09	0.10	
Test based on τ_n	0.89	0.95	0.98	0.99	0.99	
Test based on τ_n^*	1	1	1	1	1	
Test based on dcov_n	0.68	0.78	0.95	1	1	

The values in Table 1 indicate that the test based on τ_n^* is more powerful than the test based on dcov_n because of dcov_n's non-robustness property. Comparing between the tests based on τ_n and dcov_n, for small values of γ, the τ_n-based test performs better whereas for large values of γ, the test based on dcov_n performs marginally better than the test based on τ_n.

Example 2. Consider

$$\begin{aligned} f_0(x,y) &= 1 \quad \text{if} \quad (x,y) \in [0,1]^2 \\ &= 0 \quad \text{if} \quad (x,y) \notin [0,1]^2, \end{aligned}$$

and

$$\begin{aligned} k(x,y) &= 1 \quad \text{if} \quad (x,y) \in [24,25]^2 \\ &= 0 \quad \text{if} \quad (x,y) \notin [24,25]^2. \end{aligned}$$

The results are reported in Table 2.

Table 2

The results for Example 2: The asymptotic power of the different tests for different values of γ. For different values of γ, the value within each cell of the second, the third, the fourth, the sixth, the seventh and the eighth rows denote the asymptotic power of the corresponding test at 5% level of significance under contiguous alternatives.

γ	0	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006
Test based on τ_n	0.05	0.17	0.39	0.66	0.86	0.96	0.99
Test based on τ_n^*	0.05	0.25	0.41	0.75	0.91	1	1
Test based on dcov_n	0.05	0.07	0.09	0.08	0.17	0.14	0.16
γ	0.0007	0.0008	0.0009	0.001	0.05	0.10	
Test based on τ_n	0.99	1	1	1	1	1	
Test based on τ_n^*	1	1	1	1	1	1	
Test based on dcov_n	0.27	0.25	0.31	0.43	0.38	0.36	

In this example, given that the right end point of the support of F_0 is too distant from the left end point of the support of K, the test based on dcov_n does not perform well as it is a moment based procedure, i.e., non-robust against the outliers. Whereas, as expected, the test based on τ_n^* performs well since it is robust against the outliers.

Example 3. Consider $f_0(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ and $k(x, y) = \frac{1}{2\pi} e^{-\frac{(x-20)^2+(y-20)^2}{2}}$, where $(x, y) \in \mathbb{R}^2$. The results are reported in Table 3.

TABLE 3

The results for Example 3: The asymptotic power of the different tests for different values of γ. For different values of γ, the value within each cell of the second, the third, the fourth, the sixth, the seventh and the eighth rows denote the asymptotic power of the corresponding test at 5% *level of significance under contiguous alternatives.*

γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Test based on τ_n	0.05	0.08	0.15	0.23	0.33	0.44	0.56	0.68
Test based on τ_n^*	0.05	0.09	0.25	0.41	0.56	0.67	0.79	0.89
Test based on dcov_n	0.05	0.06	0.11	0.21	0.18	0.29	0.38	0.41
γ	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5
Test based on τ_n	0.77	0.85	0.91	0.95	0.97	0.98	0.99	1
Test based on τ_n^*	0.96	1	1	1	1	1	1	1
Test based on dcov_n	0.39	0.46	0.51	0.55	0.57	0.56	0.61	0.59

The figures in Table 3 also indicate that the test based on τ_n^* performs better than the test based on dcov as expected in view of the fact that $b(\beta; \tau^*(F_0))$ is bounded whereas $b(\beta; \text{dcov}(F_0))$ is unbounded. The nature of $b(\beta; \cdot)$ plays a crucial role in the power study because the distance between the location parameters of F_0 and K is large while the scatter matrices associated with F_0 and K are the same.

Example 4. Consider $f_0(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ and $k(x, y) = \frac{1}{2} \frac{1}{(1+x^2+y^2)^{\frac{3}{2}}}$ (i.e., standard bivariate Cauchy density function), where $(x, y) \in \mathbb{R}^2$. The results are reported in Table 4.

TABLE 4

The results for Example 4: The asymptotic power of the different tests for different values of γ. For different values of γ, the value within each cell of the second, the third and the fourth rows denote the asymptotic power of the corresponding test at 5% *level of significance under contiguous alternatives.*

γ	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
Test based on τ_n	0.05	0.11	0.18	0.23	0.36	0.52	0.75	0.84	0.97	0.99	1
Test based on τ_n^*	0.05	0.13	0.26	0.40	0.63	0.85	0.95	1	1	1	1
Test based on dcov_n	0.05	0.09	0.14	0.11	0.35	0.38	0.53	0.62	0.61	0.67	0.74

As expected, the figures in Table 4 indicate that the test based on τ_n^* performs best whereas the test based on dcov_n does not, since the latter lacks the robustness against the outliers generated from a heavy tailed distribution K, namely, the standard bivariate Cauchy distribution.

4. Concluding remarks

The asymptotic power study in Section 3.1 indicates that the test based on τ_n^* performs well when the null and alternative distributions are far away from each other while the test based on dcov_n does not perform well in this situation since distance covariance is not robust against outliers. On the other hand, performances of both measures are comparable when null and alternative distributions are close.

Recently, Weihs, Drton, and Leung (2016 (to appear)) provided an efficient method to compute τ_n^*. Direct computation of τ_n^* using the definition requires $O(n^4)$ operations. Similar to Christensen's 2005 idea for computing Kendall's τ, Weihs et al. observed that computing τ_n^* relies only on the relative ordering of quadruples of points. Based on this fact, they derived an algorithm to compute τ_n^* using only $O(n^2 \log(n))$ operations.

We should also point out that one can carry out two-sample tests based on τ_n^* (or dcov_n). Suppose that $\mathcal{U} = \{U_1, \ldots, U_m\}$ and $\mathcal{V} = \{V_1, \ldots, V_n\}$ are two independent sets of random variables associated with distribution functions F and G, respectively, and we want to test $H_0 : F = G$ against $H_1 : F \neq G$. We now define $(X_i, Y_i) = (U_i, 0)$ if $i = 1, \ldots, m$ and $(X_i, Y_i) = (V_{i-m}, 1)$ if $i = m+1, \ldots, m+n$. The construction of (X_i, Y_i) for $i = 1, \ldots, n+m$ implies that $X \perp\!\!\!\perp Y \Rightarrow F = G$. Note that it follows from (Bergsma and Dassios, 2014, Theorem 1)that $\tau^*(X,Y) = 0 \Leftrightarrow X \perp\!\!\!\perp Y$, which implies that $\tau^*(X,Y) = 0 \Rightarrow F = G$. In other words, the two-sample test (i.e., $H_0 : F = G$ against $H_1 : F \neq G$) is a special case of the test for independence.

One can also use τ^* to estimate the mixing proportion in the mixture distribution such as $F_{X,Y} = (1-\epsilon)F_{1X}G_{1Y} + \epsilon F_{2X}G_{2Y}$, where $\epsilon \in (0, 1/2)$ is the mixing proportion, and F_{1X}, G_{1Y}, F_{2X} and G_{2Y} are distribution functions. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. bivariate random vectors associated with $F_{1X}G_{1Y}$, and as a consequence of the product form $F_{1X}G_{1Y}$ of the joint distribution function, we have $\tau^*(X,Y) = 0$. Also, let $(X_1^*, Y_1^*), \ldots, (X_m^*, Y_m^*)$ be i.i.d. bivariate random vectors associated with $F_{2X}G_{2Y}$, and we have $\tau^*(X^*, Y^*) = 0$ in view of the product form $F_{2X}G_{2Y}$ of the joint distribution function. We now combine these n many (X, Y) and m many (X^*, Y^*) random vectors and then randomly choose n many random vectors from the combined $(n+m)$ many random vectors, which can be done in $\binom{n+m}{n}$ ways. We denote j-th set of chosen random vectors are $(x_{1j}^{**}, Y_{1j}^{**}), \ldots, (X_{nj}^{**}, Y_{nj}^{**})$, where $j = 1, \ldots, \binom{n+m}{n}$ and compute $\tau^*(X_j^{**}, Y_j^{**})$ for each $j = 1, \ldots, \binom{n+m}{n}$. Finally, in view of the structure of the mixture distribution $F_{X,Y}$, one can propose the estimate of ϵ to be

$$\hat{\epsilon}_{n,m} = \frac{\sum_{j=1}^{\binom{n+m}{n}} 1_{\{\tau^*(X_j^{**}, Y_j^{**}) > c\}}}{\binom{n+m}{n}},$$

where c is a constant, *significantly* larger than zero. The investigation of the properties of $\hat{\epsilon}_{n,m}$ is a subject for future research.

Appendix

Proof of Theorem 1. It follows from the definition of τ^* that

$$
\begin{aligned}
&\tau^*_{X,Y}\{(1-\beta)F_{X,Y}+\beta G_{X,Y}\}\\
&=Ea(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)\\
&=\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)\prod_{i=1}^{4} d\{(1-\beta)F_{X_i,Y_i}+\beta G_{X_i,Y_i}\}\\
&=(1-\beta)^4\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta(1-\beta)^3\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta(1-\beta)^3\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta(1-\beta)^3\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}H_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta^3(1-\beta)\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4})\\
&+\beta(1-\beta)^3\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})\\
&+\beta^3(1-\beta)\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})\\
&+\beta^2(1-\beta)^2\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4})
\end{aligned}
$$

$$+\beta^3(1-\beta)\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4})$$
$$+\beta^3(1-\beta)\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}H_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4})$$
$$+\beta^4\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4}).$$

Note that, since X and Y are independent, if $G_{X,Y}=\delta_{X,Y}(h,k)$, we have

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4})=0$$

and

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4})=0.$$

Also, all terms associated with either $\beta^3(1-\beta)$ or $\beta(1-\beta)^3$ converge to zero as $h,k\to\infty$. Among the terms associated with $\beta^2(1-\beta)^2$,

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4})$$

and

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}H_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})$$

converge to zero, whereas

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(G_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}H_{X_4,Y_4}),$$
$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}G_{X_3,Y_3}H_{X_4,Y_4}),$$
$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}G_{X_2,Y_2}H_{X_3,Y_3}G_{X_4,Y_4})$$

and

$$\int_{\mathbb{R}^8} a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)d(H_{X_1,Y_1}H_{X_2,Y_2}G_{X_3,Y_3}G_{X_4,Y_4})$$

converge to one as $h,k\to\infty$, in view of the definition of a. All these facts imply that $b(\beta;\tau^*)=4\beta^2(1-\beta)^2$. Also, previously mentioned in Remark 1, we will have the same expression of $b(\beta;\tau^*)=4\beta^2(1-\beta^2)$ when $h,k\to-\infty$.

Further, note that $\beta^2(1-\beta)^2$ is an increasing function for any $\beta < 1/2$, and consequently, $\beta^2(1-\beta)^2 < 1/16 \Leftrightarrow b(\beta;\tau) < 1/4$ for any $\beta < 1/2$. This completes the proof of the theorem. □

Proof of Proposition 1. For τ, considering $s(.)$ instead of $a(.)$ in the proof of Theorem 1 and arguing in the same way, we have $b(\beta;\tau) = 4\beta^2(1-\beta)^2$ and $b(\beta;\tau) < 1/4$ for any $\beta < 1/2$.

For dcov, considering $\frac{1}{4}h(.)$ instead of $a(.)$ in the expression of $\tau^*_{X,Y}\{(1-\beta)F_{X,Y} + \beta G_{X,Y}\}$, which appeared in the proof of Theorem 1, we have $\text{dcov}_{X,Y}\{(1-\beta)F_{X,Y} + \beta G_{X,Y}\}$. Note that, since X and Y are independent, if $G_{X,Y} = \delta_{X,Y}(h,k)$, we have

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(H_{X_1,Y_1} H_{X_2,Y_2} H_{X_3,Y_3} H_{X_4,Y_4}) = 0$$

and

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(G_{X_1,Y_1} G_{X_2,Y_2} G_{X_3,Y_3} G_{X_4,Y_4}) = 0.$$

Also, all terms associated with either $\beta^3(1-\beta)$ or $\beta(1-\beta)^3$ converge to zero as $h, k \to \infty$. Among the terms associated with $\beta^2(1-\beta)^2$,

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(H_{X_1,Y_1} G_{X_2,Y_2} G_{X_3,Y_3} H_{X_4,Y_4})$$

and

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(G_{X_1,Y_1} H_{X_2,Y_2} H_{X_3,Y_3} G_{X_4,Y_4})$$

converge to zero whereas

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(G_{X_1,Y_1} G_{X_2,Y_2} H_{X_3,Y_3} H_{X_4,Y_4}),$$

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(H_{X_1,Y_1} G_{X_2,Y_2} G_{X_3,Y_3} H_{X_4,Y_4}),$$

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(H_{X_1,Y_1} G_{X_2,Y_2} H_{X_3,Y_3} G_{X_4,Y_4})$$

and

$$\int_{\mathbb{R}^8} \frac{1}{4} h(X_1, X_2, X_3, X_4) h(Y_1, Y_2, Y_3, Y_4) d(H_{X_1,Y_1} H_{X_2,Y_2} G_{X_3,Y_3} G_{X_4,Y_4})$$

converge to ∞ as $h, k \to \infty$, in view of the definition of h. All these facts imply that $b(\beta; \text{dcov}) = \infty$. Similarly, here also as in the case of τ^*, the expressions of $b(\beta; \tau) = 4\beta^2(1-\beta)^2$ and $b(\beta; \text{dcov}) = \infty$ will remain the same when $h, k \to -\infty$. $\square$

To prove Theorem 2, we should state a lemma proposed by Le Cam on the asymptotic distribution of test statistics under contiguous alternatives and the result related to the asymptotic distribution of non-degenerate U-statistics.

Lemma (Le Cam's third lemma)**.** *Let $\{X_n\} \in \mathbb{R}^d$ be a sequence of random vectors, and the sequence of measures Q_n is contiguous with respect to the sequence of another probability measures P_n. If $(X_n, \log \frac{dQ_n}{dP_n})$ converges weakly to a random vector in $\mathbb{R}^{d+1}$ associated with $(d+1)$-dimensional normal distribution with the location parameter $= \begin{pmatrix} \mu \\ -\frac{\sigma^2}{2} \end{pmatrix}$ and the scatter parameter $= \begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}$ under P_n, then $\{X_n\}$ converges weakly to a random vector in $\mathbb{R}^d$ associated with d-dimensional normal distribution with the location parameter $= \mu + \tau$ and the scatter parameter $= \Sigma$ under Q_n.*

Proof. See (Van der Vaart, 2000, p. 90). $\square$

Result 1 (Asymptotic normality of non-degenerate U-statistics)**.** *For a given data $\mathcal{X} = \{x_1, \ldots, x_n\}$, let $U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \ldots < i_m} k(x_{i_1}, \ldots, x_{i_m})$ be a U-statistic of order m with kernel $k(.)$ If $\sigma_1^2 := Var_{X_1}(E_{X_2,\ldots,X_m} k(x_1, X_2, \ldots, X_m)) > 0$, then $\sqrt{n}(U_n - E[U_n])$ converges weakly to a random variable associated with the normal distribution with mean zero and variance $m^2\sigma_1^2$.*

Proof. See the proof of Theorem 1 in Lee (1990), page 76. $\square$

Proof of Theorem 2. In order to establish the contiguity of the sequence F_n relative to F_0, it is enough to show that L_n, the logarithm of the likelihood ratio, is asymptotically normal with mean $-\frac{1}{2}\sigma^2$ and variance σ^2 (see (Hajek et al., 1999, p. 253, Corollary to Le Cam's first Lemma), where σ is a positive constant. For notational convenience, we denote $\mathbf{Z} = (X, Y)$, and f_n and f_0 are the density functions of F_n and F_0, respectively. Now, we have

$$\begin{aligned} L_n &= \sum_{i=1}^{n} \log \frac{f_n(\mathbf{z}_i)}{f_0(\mathbf{z}_i)} = \sum_{i=1}^{n} \log \frac{(1-\gamma/\sqrt{n}) f_0(\mathbf{z}_i) + \gamma/\sqrt{n} k(\mathbf{z}_i)}{f_0(\mathbf{z}_i)} \\ &= \sum_{i=1}^{n} \log \left[1 + \gamma/\sqrt{n} \left\{ \frac{k(\mathbf{z}_i)}{f_0(\mathbf{z}_i)} - 1 \right\} \right] \\ &= \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} \{m(\mathbf{z}_i) - 1\} - \frac{\gamma^2}{2n} \sum_{i=1}^{n} \{m(\mathbf{z}_i) - 1\}^2 \Big/ \left\{ 1 + \frac{a_{in}\gamma \{m(\mathbf{z}_i) - 1\}}{\sqrt{n}} \right\}^2, \end{aligned}$$

where $m(\mathbf{z}_i) = \frac{k(\mathbf{z}_i)}{f_0(\mathbf{z}_i)}$, and $a_{in} \in (0,1)$ with probability 1.

Now, we define $W_n = \sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}} \{m(\mathbf{z}_i) - 1\} - \frac{\gamma^2}{2} E_{f_0} \{m(\mathbf{z}_1) - 1\}^2$. Note that by straightforward application of C.L.T., it follows that W_n is asymptotically normal with mean $-\frac{\gamma^2}{2} E_{f_0} \{m(\mathbf{z}_1) - 1\}^2$ and variance $\gamma^2 E_{f_0} \{m(\mathbf{z}_1) - 1\}^2$ since

$E_{F_0}\{\frac{k(\mathbf{z})}{f_0(\mathbf{z})}-1\}^2 < \infty$. So, in order to prove contiguity of the sequence of densities associated with H_n, it is enough to show that $|L_n - W_n| \xrightarrow{p} 0$ as $n \to \infty$.

For convenience of writing, we denote $\sigma^2 = E_{f_0}\{m(\mathbf{z}_1) - 1\}^2$, $\sigma_{1l}^2 = E_{f_0}\{m(\mathbf{z}_1) - 1\}^2 1_{\{m(\mathbf{z}_1) \leq l\}}$ and $\sigma_{2l}^2 = E_{f_0}\{m(\mathbf{z}_1) - l\}^2 1_{\{m(\mathbf{z}_1) > l\}}$, where $l > 0$ is a constant. So, we have

$$|L_n - W_n| \leq \left| T_{1n} - \frac{\gamma^2 \sigma_{1l}^2}{2} \right| + \left| T_{2n} - \frac{\gamma^2 \sigma_{2l}^2}{2} \right|,$$

where T_{1n} and T_{2n} are given by $T_{1n} = \sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2}{2n} / [1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}}]^2 \times 1_{\{m(\mathbf{z}_i) \leq l\}}$ and $T_{2n} = \sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2}{2n} / [1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}}]^2 1_{\{m(\mathbf{z}_i) > l\}}$. Now, for a fixed $\epsilon > 0$, we choose l_0 sufficiently large such that $\gamma^2 \sigma_{2l_0}^2 < \epsilon$ and $l_0 > 1$. Then,

$$\begin{aligned} &P\left[\left| T_{2n} - \frac{\gamma^2 \sigma_{2l_0}^2}{2} \right| > \epsilon/2 \right] \\ &\quad = P\left[T_{2n} > \frac{\gamma^2 \sigma_{2l_0}^2}{2} + \frac{\epsilon}{2} \right] + P\left[T_{2n} < \frac{\gamma^2 \sigma_{2l_0}^2}{2} - \frac{\epsilon}{2} \right] \\ &\quad = P\left[T_{2n} > \frac{\gamma^2 \sigma_{2l_0}^2}{2} + \frac{\epsilon}{2} \right] + 0 \ (\text{since } \tfrac{\gamma^2 \sigma_{2l_0}^2}{2} - \tfrac{\epsilon}{2} < 0) \\ &\quad \leq P\left[\sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i) - 1\}^2}{2n} 1_{\{m(\mathbf{z}_i) > l_0\}} > \frac{\gamma^2 \sigma_{2l_0}^2}{2} + \frac{\epsilon}{2} \right] \to 0 \text{ as } n \to \infty. \end{aligned}$$

The last implication follows from the fact that $\sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2}{2n} 1_{\{m(\mathbf{z}_i) > l_0\}} \xrightarrow{p} \frac{\gamma^2 \sigma_{2l_0}^2}{2}$.

Now, we fix $0 < \eta < 1$ on the event $\{m(\mathbf{z}_i) \leq l_0\}$, and hence, we have $1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}} \leq 1 + \frac{\gamma(l_0-1)}{\sqrt{n}} < 1 + \eta$ for all $n \geq N$ (say). Also, since $m(\mathbf{z}) \geq 0$, $1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}} \geq 1 - \frac{a_{in}\gamma}{\sqrt{n}} > 1 - \eta$ for all $n \geq N$. Next, we define $V_{1n} = \sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2 1_{\{m(\mathbf{z}_i) \leq l_0\}}}{2n(1+\eta)^2}$ and $V_{2n} = \sum_{i=1}^n \frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2 1_{\{m(\mathbf{z}_i) \leq l_0\}}}{2n(1-\eta)^2}$. The aforementioned facts imply that $T_{1n} \in (V_{1n}, V_{2n})$ for all $n \geq N$, $V_{1n} \xrightarrow{p} \frac{\gamma^2 \sigma_{1l_0}^2}{2(1+\eta)^2}$ and $V_{2n} \xrightarrow{p} \frac{\gamma^2 \sigma_{1l_0}^2}{2(1-\eta)^2}$. Hence, we have

$$\begin{aligned} P\left[\left| T_{1n} - \frac{\gamma^2 \sigma_{1l_0}^2}{2} \right| > \epsilon/2 \right] &\leq P\left[\left| V_{1n} - \frac{\gamma^2 \sigma_{1l_0}^2}{2(1+\eta)^2} \right| > \epsilon/2 - \frac{\gamma^2 \sigma_{1l_0}^2}{2} + \frac{\gamma^2 \sigma_{1l_0}^2}{2(1+\eta)^2} \right] \\ &\quad + P\left[\left| V_{2n} - \frac{\gamma^2 \sigma_{1l_0}^2}{2(1-\eta)^2} \right| > \epsilon/2 + \frac{\gamma^2 \sigma_{1l_0}^2}{2} + \frac{\gamma^2 \sigma_{1l_0}^2}{2(1-\eta)^2} \right]. \end{aligned}$$

Now, we choose $\eta > 0$ so small such that $\epsilon/2 - \frac{\gamma^2 \sigma_{1l_0}^2}{2} + \frac{\gamma^2 \sigma_{1l_0}^2}{2(1+\eta)^2} > 0$ and $\epsilon/2 + \frac{\gamma^2 \sigma_{ll_0}^2}{2} + \frac{\gamma^2 \sigma_{1l_0}^2}{2(1-\eta)^2} > 0$. Thus, we have $T_{1n} - \frac{\gamma^2 \sigma_{1l_0}^2}{2} \xrightarrow{p} 0$, and consequently,

$L_n - W_n \xrightarrow{p} 0$, which ensures the contiguity of the sequence of densities associated with H_n.

Next, since τ_n is a non-degenerate U-statistics (see Lee (1990), p. 14–15) as $E_{X_1,Y_1}[E\text{sign}\{(X_1 - X_2)(Y_1 - Y_2)|(X_1, Y_1)\}]^2 = \int_{\infty}^{\infty}\int_{\infty}^{\infty}[2\int_x^{\infty}\int_y^{\infty} f_0(u,v)dudv + 2\int_{-\infty}^{x}\int_{-\infty}^{y} f_0(u,v)dudv - 1]^2 f_0(x,y)dxdy > 0$, it follows from Result 1 (see also Lee (1990), Theorem 1, p. 76) that $\sqrt{n}\{(\tau_n - \tau)$ converges weakly to a random variable associated with a normal distribution with zero mean and variance σ_1^2, where the expression of σ_1^2 is provided later. Further, in view of expansion of L_n, we have the asymptotic normality of the all possible linear combination of $\sqrt{n}(\tau_n - \tau)$ and L_n under H_0. This fact implies that the joint distribution of $\sqrt{n}\{(\tau_n - \tau), L_n/\sqrt{n}\}$ is asymptotically bivariate normal distribution under H_0. Hence, one can apply Le Cam's third lemma (see the statement of this lemma before this proof and also see Hajek et al. (1999)) to establish the asymptotic distribution of $\sqrt{n}\{(\tau_n - \tau)$ under H_n. It is here appropriate to note that under H_n also, $\sqrt{n}\{(\tau_n - \tau)$ weakly converges to a Gaussian random variable as under H_0 but a location shift occurs in the expression of the location parameter of the Gaussian distribution. Le Cam's third lemma indicates that the location shift is essentially the asymptotic covariance between $\sqrt{n}(\tau_n - \tau)$ and L_n. Now, the asymptotic covariance between $\sqrt{n}(\tau_n - \tau)$ and L_n is

$$\frac{2\gamma}{n}E_{f_0}\left[\sum_{i=1}^{n}\psi_1(X_i,Y_i)\times\left\{\frac{k(\mathbf{z}_i)}{f_0(\mathbf{z}_i)} - 1\right\}\right],$$

$$\text{where } \psi_1(X_i,Y_i) = E\text{sign}\{(X_i - X_j)(Y_i - Y_j)|(X_i,Y_i)\}$$

$$= 2\gamma E_k\psi_1(X,Y) - 2\gamma E_{f_0}\psi_1(X,Y) = 2\gamma E_k\psi_1(X,Y)$$

$$\text{since } 2E_{f_0}\psi_1(X,Y) = 0$$

$$= 2\gamma\int_{\infty}^{\infty}\int_{\infty}^{\infty}\left[2\int_{x}^{\infty}\int_{y}^{\infty} f_0(u,v)dudv + 2\int_{-\infty}^{x}\int_{-\infty}^{y} f_0(u,v)dudv - 1\right]k(x,y)dxdy.$$

Hence, Le Cam's third lemma leads to the conclusion that under contiguous alternatives H_n, $\sqrt{n}(\tau_n - \tau)$ converges weakly to a Gaussian distribution with mean

$$\mu_1 = 2\gamma\int_{\infty}^{\infty}\int_{\infty}^{\infty}\left[2\int_{x}^{\infty}\int_{y}^{\infty} f_0(u,v)dudv + 2\int_{-\infty}^{x}\int_{-\infty}^{y} f_0(u,v)dudv - 1\right]k(x,y)dxdy$$

and variance

$$\sigma_1^2 = 4\int_{\infty}^{\infty}\int_{\infty}^{\infty}\left[2\int_{x}^{\infty}\int_{y}^{\infty} f(u,v)dudv + 2\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv - 1\right]^2 f(x,y)dxdy.$$

This completes the proof. □

Proof of Theorem 3. We first note that τ_n^* is a U-statistic having a degeneracy of order 1, which follows from the following. Note that

$$\begin{aligned} & E_{f_0}\{a(X_1,X_2,X_3,X_4)a(Y_1,Y_2,Y_3,Y_4)|(X_1,Y_1)\} \\ = \; & E_{f_0}\{a(X_1,X_2,X_3,X_4)|X_1\}E_{f_0}\{a(Y_1,Y_2,Y_3,Y_4)|Y_1\} \text{ since } F_0 = G_X H_Y. \end{aligned}$$

So, in order to establish that fact, it is now enough to show that $E\{a(X_1,X_2,X_3,X_4)|(X_1=x_1)\} = 0$ for all x_1.

Hence, we consider

$$\begin{aligned} Ea(x_1,X_2,X_3,X_4) &= E\text{sign}(|x_1-X_2|+|X_3-X_4|-|x_1-X_3|-|X_2-X_4|) \\ &= P[|x_1-X_2|+|X_3-X_4|-|x_1-X_3|-|X_2-X_4|>0] \\ &- P[|x_1-X_2|+|X_3-X_4|-|x_1-X_3|-|X_2-X_4|<0] \\ &= P[|x_1-X_2|+|X_3-X_4|>|x_1-X_3|+|X_2-X_4|] \\ &- P[|x_1-X_2|+|X_3-X_4|<|x_1-X_3|+|X_2-X_4|] \\ &= \frac{1}{2}-\frac{1}{2} = 0 \text{ for all } x_1. \end{aligned}$$

The last step follows from the fact that X_2, X_3 and X_4 are i.i.d. random variables. Also, it is easy to see that $E\{a(X_1,X_2,X_3,X_4)|(X_1=x_1,X_2=x_2)\} \neq 0$ for some x_1 and x_2. Hence, it is now established that τ_n^* is a U-statistic having a degeneracy of order 1.

Further, note that the densities (denoted as q_n) associated with H_n is dominated by the density (denote it as p_0) associated with H_0 with Radon-Nikodym derivative $\frac{dq_n}{dp_0} = 1+n^{-\frac{1}{2}}h_n$, where $h_n = \gamma(\frac{k}{f_0}-1) \in L_2(p_0)$ since $E_{f_0}(\frac{k}{f_0}-1)^2 < \infty$, which is asserted in the statement of Theorem 2. Hence, q_n and p_0 satisfy the assumptions stated in Theorem 1 in Gregory (1977), which concludes that $n(\tau_n^*-\tau^*)$ converges weakly to $\sum_{i=1}^{\infty}\lambda_i\{(Z_i+a_i)^2-1\}$ under H_n, where λ_i, Z_i and a_i are as defined in the statement of the theorem. This completes the proof. □

Proof of Theorem 4. We first note that dcov_n is a U-statistic having a degeneracy of order 1, which follows from the following. Note that

$$\begin{aligned} & \frac{1}{4}E_{f_0}\{h(X_1,X_2,X_3,X_4)h(Y_1,Y_2,Y_3,Y_4)|(X_1,Y_1)\} \\ = \; & \frac{1}{4}E_{f_0}\{h(X_1,X_2,X_3,X_4)|X_1\}E_{f_0}\{h(Y_1,Y_2,Y_3,Y_4)|Y_1\} \text{ since } F_0 = G_X H_Y. \end{aligned}$$

So, in order to establish that fact, it is now enough to show that $E\{h(X_1,X_2,X_3,X_4)|(X_1=x_1)\} = 0$ for all x_1.

Hence, we consider $Eh(x_1,X_2,X_3,X_4) = \int_{\mathbb{R}^3}\{|x_1-X_2|+|X_3-X_4|-|x_1-X_3|-|X_2-X_4|\}\prod_{i=2}^{4}dG_{X_i} = 0$ for all x_1 in view of the fact that X_2, X_3 and X_4 are i.i.d. random variables. In addition, it is easy to see that $E\{h(X_1,X_2,X_3,X_4)|(X_1=x_1,X_2=x_2)\} \neq 0$ for some x_1 and x_2. Hence, it is now established that dcov_n is a U-statistic having a degeneracy of order 1.

Next, arguing in a similar way as in the last paragraph of the proof of Theorem 3, we have $n(\text{dcov}_n - \text{dcov})$ converging weakly to $\sum_{i=1}^{\infty} \lambda_i^* \{(Z_i^* + a_i^*)^2 - 1\}$ under H_n, where λ_i^*, Z_i^* and a_i^* are as defined in the statement of the theorem. This completes the proof. □

Acknowledgements

The authors would like to thank the editor Professor George Michailidis, an anonymous associate editor and two anonymous referees for several useful comments, which improved the article significantly. The work was started when the first author was visiting the Department of Statistics, London School of Economics and Political Science, and he would like to thank them for their hospitality. This work is partially supported by the first author's CPDA grant (Project Name: IITK /MATH/20130171) from the IIT Kanpur. The first author is also thankful to Ms. Asmita Gupta for helping him in editing the first draft of the article and Professor Hira Koul for some helpful comments.

References

Babu, G. J., & Rao, C. R. (1988). Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis*, *27*(1), 15–23. MR0971169

Banerjee, M. (2005). Likelihood ratio tests under local alternatives in regular semiparametric models. *Statistica Sinica*, *15*(3), 635–644. MR2233903

Behnen, K. (1971). Asymptotic optimality and ARE of certain rank-order tests under contiguity. *The Annals of Mathematical Statistics*, 325–329. MR0275593

Behnen, K., & Neuhaus, G. (1975). A central limit theorem under contiguous alternatives. *The Annals of Statistics*, *3*(6), 1349–1353. MR0388611

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

Bergsma, W. P. (2006). A new correlation coefficient, its orthogonal decomposition, and associated tests of independence. *arXiv:math/0604627v1 [math.ST]*.

Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, *32*, 485–498. MR0125690

Christensen, D. (2005). Fast algorithms for the calculation of Kendall's τ. *Computational Statistics*, *20*(1), 51–62. MR2162534

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference.* Springer. MR2681063

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *The Annals of Statistics*, 110–123. MR0433669

HAJEK, J., SIDAK, Z., & SEN, P. K. (1999). *Theory of rank tests.* Academic Press. MR1680991

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, *19*, 293–325. MR0026294

HUBER, P. J. (2011). *Robust statistics (2^{nd} edition).* Springer. MR2743408

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

LEE, A. (1990). *U*-statistics. *Theory and Practice*, Marcel Dekker, New York. MR1075417

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hung.*, *10*, 441–451. MR0115203

SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics.* New York: Wiley-Blackwell. MR0595165

SPEARMAN, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101.

SZÉKELY, G. J., RIZZO, M. L., & BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. MR2382665

VAN DER VAART, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge University Press. MR1652247

WEIHS, L., DRTON, M., & LEUNG, D. (2016 (to appear)). Efficient computation of the Bergsma-Dassios sign covariance. *Computational Statistics*.