Biometrika (2022), **109**, 2, *pp.* 317–333
doi: 10.1093/biomet/asab028
Advance Access publication 6 October 2021

On the power of Chatterjee's rank correlation

By H. SHI

Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195, U.S.A.

hongshi@uw.edu

M. DRTON

Department of Mathematics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching b. München, Germany

mathias.drton@tum.de

AND F. HAN

Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195, U.S.A.

fanghan@uw.edu

Summary

Chatterjee (2021) introduced a simple new rank correlation coefficient that has attracted much attention recently. The coefficient has the unusual appeal that it not only estimates a population quantity first proposed by Dette et al. (2013) that is zero if and only if the underlying pair of random variables is independent, but also is asymptotically normal under independence. This paper compares Chatterjee's new correlation coefficient with three established rank correlations that also facilitate consistent tests of independence, namely Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^*. We compare the computational efficiency of these rank correlation coefficients in light of recent advances, and investigate their power against local rotation and mixture alternatives. Our main results show that Chatterjee's coefficient is unfortunately rate-suboptimal compared to D, R and τ^*. The situation is more subtle for a related earlier estimator of Dette et al. (2013). These results favour D, R and τ^* over Chatterjee's new correlation coefficient for the purpose of testing independence.

Some key words: Dependence measure; Independence test; Le Cam's third lemma; Rank correlation; Rate-optimality.

1. Introduction

Let $X^{(1)}$ and $X^{(2)}$ be two real-valued random variables defined on a common probability space. We are concerned with testing the null hypothesis

$$H_0 : X^{(1)} \text{ and } X^{(2)} \text{ are independent,}$$

based on a sample from the joint distribution of $(X^{(1)}, X^{(2)})$. This classical problem has seen revived interest in recent years as independence tests constitute a key component in modern statistical methodology such as methods for causal discovery (e.g. Maathuis et al., 2019, § 18.6.3).

The problem of testing independence has been examined from a number of different perspectives; see, for example, the work of Kim et al. (2020), Albert et al. (2021) and Berrett et al. (2021) and and the references therein. In this paper, our focus will be on testing H_0 via rank correlations that measure ordinal association. Rank correlations are particularly attractive for continuous distributions for which they are distribution-free under H_0. Early proposals of rank correlations include the widely used ρ of Spearman (1904) and τ of Kendall (1938), as well as the footrule of Spearman (1906), the γ of Gini (1914) and the β of Blomqvist (1950). Unfortunately, all five of these rank correlations fail to give a consistent test of independence. Indeed, each correlation coefficient consistently estimates a population correlation measure that takes the same value under H_0 and certain fixed alternatives to H_0. This behaviour leads to trivial power at such alternatives.

In order to arrive at a consistent test of independence, Hoeffding (1948) proposed a correlation measure that, for absolutely continuous bivariate distributions, vanishes if and only if H_0 holds. Blum et al. (1961) considered a modification that is consistent against all dependent bivariate alternatives (cf. Hoeffding, 1940). Bergsma & Dassios (2014) proposed a new test of independence and showed its consistency for bivariate distributions that are discrete, absolutely continuous, or a mixture of the two types. As pointed out by Drton et al. (2020), mere continuity of the marginal distribution functions is sufficient for consistency of their test. This follows from a relation discovered by Yanagimoto (1970), who implicitly considered the correlation of Bergsma & Dassios (2014) when proving a conjecture of Hoeffding (1948).

All three aforementioned correlation measures admit natural efficient estimators in the form of U-statistics that depend only on ranks. However, in each case, the U-statistic is degenerate and has a nonnormal asymptotic distribution under H_0. In light of this fact, it is interesting that Dette et al. (2013) were able to construct a consistent correlation measure ξ which can also detect perfect functional dependence (see also Gamboa et al., 2018); and in a recent paper that has received much attention, Chatterjee (2021) introduced a very simple rank correlation, with no tuning parameter involved, that surprisingly estimates ξ and has an asymptotically normal null distribution.

In this paper we compare Chatterjee's and Dette–Siburg–Stoimenov's rank correlation coefficients with three obvious competitors: the D of Hoeffding (1948), the R of Blum et al. (1961) and the τ^* of Bergsma & Dassios (2014). Our comparison considers three criteria:

(i) *Statistical consistency of the independence test.* A correlation measure μ assigns to each joint distribution of $(X^{(1)}, X^{(2)})$ a real number $\mu(X^{(1)}, X^{(2)})$. Such a correlation measure is consistent within a family of distributions $\mathcal{F}$ if for all pairs $(X^{(1)}, X^{(2)})$ with joint distribution in $\mathcal{F}$, $\mu(X^{(1)}, X^{(2)}) = 0$ if and only if $X^{(1)}$ is independent of $X^{(2)}$. Correlation measures that are consistent within a large nonparametric family are able to detect nonlinear, nonmonotone relationships and facilitate consistent tests of independence. If a correlation measure μ is consistent, then the consistency of tests of independence based on an estimator μ_n of μ is guaranteed by the consistency of that estimator.

(ii) *Computational efficiency.* Computation of ranks requires $O(n \log n)$ time. With a view towards large-scale applications, we prioritize rank correlation coefficients that are computable without much additional effort, that is, also in $O(n \log n)$ time. This is easily seen to be the case for Chatterjee's coefficient, but, as we shall survey in § 2, recent advances have clarified that D, R and τ^* can be computed with a similar level of efficiency.

(iii) *Statistical efficiency of the independence test.* Our final criterion is optimal efficiency in the statistical sense (Nikitin, 1995, § 5.4). To assess this, we use different local alternatives inspired by the work of Konijn (1956) and Farlie (1960, 1961); the latter type of alternatives was further developed in Dhar et al. (2016). We then refer to an independence test as rate-optimal or rate-suboptimal against a family of local alternatives according to whether or not the test achieves the detection boundary within this family.

The main contribution of this paper pertains to statistical efficiency. Chatterjee's derivation of asymptotic normality for his rank correlation coefficient relies on a reformulation of his statistic and the use of a type of permutation central limit theorem that was established in Chao et al. (1993). We have found that a direct use of this technique to analyse the local power is hard. In recent related work we have been able to overcome a similar issue in a related multivariate setting (Deb & Sen, 2019; Shi et al., 2021) by developing a suitable Hájek representation theory (Shi et al., 2020). Following the same philosophy here, we construct a particular form of the projected statistic introduced in Angus (1995) to provide an alternative proof of Theorem 2.1 in Chatterjee (2021) that gives an asymptotic representation. Integrating the representation into Le Cam's third lemma and employing further a version of the conditional multiplier central limit theorem (cf. van der Vaart & Wellner, 1996, § 2.9), we can then show that the test based on Chatterjee's rank correlation coefficient is in fact rate-suboptimal against the two local alternative families under consideration; recall point (iii) above. Our theoretical analysis thus echos Chatterjee's empirical observation that his test of independence can suffer from low power; see Remark 7 below. In contrast, the tests based on the more established coefficients D, R and τ^* are all rate-optimal for all local alternative families considered. We therefore regard the latter as being more suitable for testing independence than Chatterjee's test. On the other hand, the test based on Dette–Siburg–Stoimenov's coefficient is empirically observed to have nontrivial power against certain alternatives in finite-sample simulations. A theoretical study of this phenomenon, however, has to be left to future work because of the technical difficulties involved. The proofs of our theoretical results, including the details of examples, are given in the Supplementary Material.

As we were completing this study, we became aware of independent work by Cao & Bickel (2020), who conducted a similar local power analysis for Chatterjee's correlation coefficient and presented a result that is similar to our Theorem 1, claim (12). The local alternatives considered in their paper are, however, different from ours. In addition, the two papers differ in their focus. The work of Cao & Bickel concentrates on correlation measures that are 1 if and only if one variable is a shape-restricted function of the other variable, whereas our interest is in comparing consistent tests of independence.

2. Rank correlations and independence tests

2.1. *The rank correlations considered and their computation*

When considering correlations, we will use the term correlation measure to refer to population quantities, which we write using Greek or Latin letters. The term correlation coefficient is reserved for sample quantities, which are written with an added subscript n. The symbol F denotes a joint bivariate distribution function for the pair of random variables $(X^{(1)}, X^{(2)})$ under consideration, and F_1 and F_2 are the respective marginal distribution functions. Throughout, $(X_1^{(1)}, X_1^{(2)}), \ldots, (X_n^{(1)}, X_n^{(2)})$ is a sample consisting of n independent copies of $(X^{(1)}, X^{(2)})$.

We now introduce in precise terms the five types of rank correlation considered in this paper. We begin by specifying the correlation measure and coefficients from Chatterjee (2021) and Dette et al. (2013). To this end, let $(X_{[1]}^{(1)}, X_{[1]}^{(2)}), \ldots, (X_{[n]}^{(1)}, X_{[n]}^{(2)})$ be a rearrangement of the sample such that $X^{(1)}_{[1]} \leqslant \cdots \leqslant X^{(1)}_{[n]}$, with ties, if they exist, broken at random. Define

$$r_{[i]} = \sum_{j=1}^{n} I\big(X^{(2)}_{[j]} \leqslant X^{(2)}_{[i]}\big), \tag{1}$$

where $I(\cdot)$ represents the indicator function, and $\ell_{[i]} = \sum_{j=1}^{n} I(X^{(2)}_{[j]} \geqslant X^{(2)}_{[i]})$. We emphasize that if F_2 is continuous, then there are almost surely no ties among $X^{(2)}_1, \ldots, X^{(2)}_n$, in which case $r_{[i]}$ is simply the rank of $X^{(2)}_{[i]}$ among $X^{(2)}_{[1]}, \ldots, X^{(2)}_{[n]}$.

DEFINITION 1. *The correlation coefficient of Chatterjee (2021) is*

$$\xi_n = 1 - \frac{n\sum_{i=1}^{n-1} |r_{[i+1]} - r_{[i]}|}{2\sum_{i=1}^{n} \ell_{[i]}(n - \ell_{[i]})}. \tag{2}$$

If there are no ties among $X^{(2)}_1, \ldots, X^{(2)}_n$, then

$$\xi_n = 1 - \frac{3\sum_{i=1}^{n-1} |r_{[i+1]} - r_{[i]}|}{n^2 - 1}.$$

Chatterjee (2021) proved that ξ_n estimates the correlation measure

$$\xi = \frac{\int \mathrm{var}[E\{I(X^{(2)} \geqslant x) \mid X^{(1)}\}]\, \mathrm{d}F_2(x)}{\int \mathrm{var}\{I(X^{(2)} \geqslant x)\}\, \mathrm{d}F_2(x)}.$$

This measure was in fact first proposed by Dette et al. (2013); cf. $r(X, Y)$ in their Theorem 2. We therefore refer to ξ as Dette–Siburg–Stoimenov's rank correlation measure.

We remark that ξ was also considered by Gamboa et al. (2018); see the Cramér–von Mises index $S^{v}_{2,\mathrm{CVM}}$ before their Properties 3.2. For estimation of ξ, Dette et al. (2013) proposed the following coefficient, denoted by $\hat{r}_n$, in their equation (15).

DEFINITION 2. *Let K be a symmetric and twice continuously differentiable kernel with compact support, and let $\bar{K}(x) = \int_{-\infty}^{x} K(t)\, \mathrm{d}t$. Let $h_1, h_2 > 0$ be bandwidths chosen such that they tend to zero with $nh_1^3 \to \infty$, $nh_1^4 \to 0$, $nh_2^4 \to 0$, $nh_1h_2 \to \infty$ as $n \to \infty$. Define*

$$\zeta_n\big(u^{(1)}, u^{(2)}\big) = \frac{1}{nh_1} \sum_{i=1}^{n} K\left(\frac{u^{(1)} - i/n}{h_1}\right) \bar{K}\left(\frac{u^{(2)} - r_{[i]}/n}{h_2}\right)$$

with $r_{[i]}$ as in (1). Then Dette–Siburg–Stoimenov's correlation coefficient is

$$\xi^*_n = 6 \int_0^1 \int_0^1 \big\{\zeta_n\big(u^{(1)}, u^{(2)}\big)\big\}^2\, \mathrm{d}u^{(1)}\, \mathrm{d}u^{(2)} - 2.$$

Next we introduce two classical rank correlations, those of Hoeffding (1948) and Blum et al. (1961), which both assess dependence in a very intuitive way by integrating squared deviations between the joint distribution function and the product of the marginal distribution functions.

Definition 3. *Hoeffding's correlation measure is defined as*

$$D = \int \left\{F\left(x^{(1)}, x^{(2)}\right) - F_1\left(x^{(1)}\right)F_2\left(x^{(2)}\right)\right\}^2 \mathrm{d}F\left(x^{(1)}, x^{(2)}\right).$$

It is unbiasedly estimated by the correlation coefficient

$$\begin{aligned} D_n = &\frac{1}{n(n-1)\cdots(n-4)} \\ &\times \sum_{i_1 \neq \ldots \neq i_5} \frac{1}{4}\left[\left\{I\left(X_{i_1}^{(1)} \leqslant X_{i_5}^{(1)}\right) - I\left(X_{i_2}^{(1)} \leqslant X_{i_5}^{(1)}\right)\right\}\left\{I\left(X_{i_3}^{(1)} \leqslant X_{i_5}^{(1)}\right) - I\left(X_{i_4}^{(1)} \leqslant X_{i_5}^{(1)}\right)\right\}\right] \\ &\times \left[\left\{I\left(X_{i_1}^{(2)} \leqslant X_{i_5}^{(2)}\right) - I\left(X_{i_2}^{(2)} \leqslant X_{i_5}^{(2)}\right)\right\}\left\{I\left(X_{i_3}^{(2)} \leqslant X_{i_5}^{(2)}\right) - I\left(X_{i_4}^{(2)} \leqslant X_{i_5}^{(2)}\right)\right\}\right], \end{aligned} \tag{3}$$

which is a rank-based U-statistic of order 5.

Definition 4. *Blum–Kiefer–Rosenblatt's correlation measure is defined as*

$$R = \int \left\{F\left(x^{(1)}, x^{(2)}\right) - F_1\left(x^{(1)}\right)F_2\left(x^{(2)}\right)\right\}^2 \mathrm{d}F_1\left(x^{(1)}\right) \mathrm{d}F_2\left(x^{(2)}\right).$$

It is unbiasedly estimated by Blum–Kiefer–Rosenblatt's correlation coefficient

$$\begin{aligned} R_n = &\frac{1}{n(n-1)\cdots(n-5)} \\ &\times \sum_{i_1 \neq \ldots \neq i_6} \frac{1}{4}\left[\left\{I\left(X_{i_1}^{(1)} \leqslant X_{i_5}^{(1)}\right) - I\left(X_{i_2}^{(1)} \leqslant X_{i_5}^{(1)}\right)\right\}\left\{I\left(X_{i_3}^{(1)} \leqslant X_{i_5}^{(1)}\right) - I\left(X_{i_4}^{(1)} \leqslant X_{i_5}^{(1)}\right)\right\}\right] \\ &\times \left[\left\{I\left(X_{i_1}^{(2)} \leqslant X_{i_6}^{(2)}\right) - I\left(X_{i_2}^{(2)} \leqslant X_{i_6}^{(2)}\right)\right\}\left\{I\left(X_{i_3}^{(2)} \leqslant X_{i_6}^{(2)}\right) - I\left(X_{i_4}^{(2)} \leqslant X_{i_6}^{(2)}\right)\right\}\right], \end{aligned} \tag{4}$$

which is a rank-based U-statistic of order 6.

More recently, Bergsma & Dassios (2014) introduced the following rank correlation, which is connected to work by Yanagimoto (1970). We refer the reader to Bergsma & Dassios (2014) for a motivation via the concordance/discordance of four-point patterns and connections to Kendall's tau.

Definition 5. *Write* $I(x_1, x_2 < x_3, x_4) = I(\max\{x_1, x_2\} < \min\{x_3, x_4\})$. *Define Bergsma–Dassios–Yanagimoto's correlation measure by*

$$\begin{aligned} \tau^* = &\, 4\,\mathrm{pr}\left(X_1^{(1)}, X_3^{(1)} < X_2^{(1)}, X_4^{(1)},\ X_1^{(2)}, X_3^{(2)} < X_2^{(2)}, X_4^{(2)}\right) \\ &+ 4\,\mathrm{pr}\left(X_1^{(1)}, X_3^{(1)} < X_2^{(1)}, X_4^{(1)},\ X_2^{(2)}, X_4^{(2)} < X_1^{(2)}, X_3^{(2)}\right) \\ &- 8\,\mathrm{pr}\left(X_1^{(1)}, X_3^{(1)} < X_2^{(1)}, X_4^{(1)},\ X_1^{(2)}, X_4^{(2)} < X_2^{(2)}, X_3^{(2)}\right). \end{aligned}$$

It is unbiasedly estimated by a U-statistic of order 4*, namely Bergsma–Dassios–Yanagimoto's correlation coefficient*

$$\begin{aligned}\tau_n^* = {} & \frac{1}{n(n-1)(n-2)(n-3)} \\ & \times \sum_{i_1 \neq \ldots \neq i_4} \left\{ I\left(X_{i_1}^{(1)}, X_{i_3}^{(1)} < X_{i_2}^{(1)}, X_{i_4}^{(1)}\right) + I\left(X_{i_2}^{(1)}, X_{i_4}^{(1)} < X_{i_1}^{(1)}, X_{i_3}^{(1)}\right) \right. \\ & \left. - I\left(X_{i_1}^{(1)}, X_{i_4}^{(1)} < X_{i_2}^{(1)}, X_{i_3}^{(1)}\right) - I\left(X_{i_2}^{(1)}, X_{i_3}^{(1)} < X_{i_1}^{(1)}, X_{i_4}^{(1)}\right) \right\} \\ & \times \left\{ I\left(X_{i_1}^{(2)}, X_{i_3}^{(2)} < X_{i_2}^{(2)}, X_{i_4}^{(2)}\right) + I\left(X_{i_2}^{(2)}, X_{i_4}^{(2)} < X_{i_1}^{(2)}, X_{i_3}^{(2)}\right) \right. \\ & \left. - I\left(X_{i_1}^{(2)}, X_{i_4}^{(2)} < X_{i_2}^{(2)}, X_{i_3}^{(2)}\right) - I\left(X_{i_2}^{(2)}, X_{i_3}^{(2)} < X_{i_1}^{(2)}, X_{i_4}^{(2)}\right) \right\}. \end{aligned} \tag{5}$$

Remark 1 (*Relation between* D_n, R_n *and* τ_n^*). As conveyed by equation (6.1) in Drton et al. (2020), as long as $n \geqslant 6$ and there are no ties in the data, one has $12D_n + 24R_n = \tau_n^*$. Consequently, $12D + 24R = \tau^*$ given continuity, but not necessarily absolute continuity, of F; cf. Yanagimoto (1970, p. 62).

At first sight the computation of the different correlation coefficients appears to be of very different complexity. However, this is not the case owing to recent developments, which yield nearly linear computation time for all coefficients except ξ_n^*.

PROPOSITION 1 (COMPUTATIONAL EFFICIENCY). *If the data contain no ties, then* ξ_n*,* D_n*,* R_n *and* τ_n^* *can all be computed in* $O(n \log n)$ *time.*

Proof. From its simple form it is evident that ξ_n can be computed in $O(n \log n)$ time (Chatterjee, 2021, Remark 4). The result about D_n is due to Hoeffding (1948, § 5); see also Weihs et al. (2018, p. 557). The claim about τ_n^* is based on recent new methods due to Even-Zohar & Leng (2021, Corollary 4) and Even-Zohar (2020b, Theorem 6.1); for an implementation see Even-Zohar (2020a). The claim about R_n then follows from the relation given in Remark 1. □

Remark 2 (*Computation of* ξ_n^*). The definition of ξ_n^* involves an integral over the unit square $[0, 1]^2$. How quickly the integral can be computed depends on smoothness properties of the kernel considered and the choice of bandwidth. Chatterjee (2021, Remark 5) suggests a time complexity of $O(n^{5/3})$. Indeed, for a symmetric and four-times continuously differentiable kernel K with compact support, there is a choice of bandwidths h_1 and h_2 that satisfies the requirements of Definition 2 and for which ξ_n^* can be approximated with an absolute error of order $o(n^{-1/2})$ in $O(n^{5/3})$ time.

To accomplish this we may choose $h_1 = h_2 = n^{-1/4-\epsilon}$ for small $\epsilon > 0$ and apply Simpson's rule to the two-dimensional integral in the definition of ξ_n^*. By the assumptions on K, the function ζ_n^2 has continuous and compactly supported fourth partial derivatives that are bounded by a constant multiple of h_1^{-5}. The error of Simpson's rule applied with a grid of M^2 points in $[0, 1]^2$ is then $O(h_1^{-5}/M^4)$. With $M^2 = O(h_1^{-5/2} n^{1/4+\epsilon/2}) = O(n^{7/8+3\epsilon})$, this error becomes $O(n^{-1/2-\epsilon}) = o(n^{-1/2})$. Because of the compact support of K, one evaluation of ζ_n requires $O(nh_1)$ operations. The overall computational time is thus $O(nh_1M^2) = O(n^{13/8+2\epsilon})$, which is $O(n^{5/3})$ as long as $\epsilon \leqslant 1/48$.

Remark 3 (*Computation with ties*). When the data can be treated as being generated from a continuous distribution, but featuring a small number of ties due to rounding, then ad-hoc breaking of ties poses little problem. In contrast, if ties arise from some discontinuity of the data-generating distribution, then the situation is more subtle. In this case, Chatterjee's ξ_n is to be computed in the form (2), but the computational time clearly remains $O(n \log n)$. In contrast, ξ_n^* is no longer a suitable estimator of ξ. Hoeffding's formulas for D_n continue to apply with ties, keeping the computation at $O(n \log n)$; however, as we shall emphasize in § 4, the estimated D may lose some of its appeal. Bergsma–Dassios–Yanagimoto's τ_n^* is suitable also for discrete data, but the available implementations that explicitly account for data with ties (Weihs, 2019) are based on the $O(n^2 \log n)$ algorithm of Weihs et al. (2016, § 3) or the slightly more memory-intensive, but faster $O(n^2)$ algorithm of Heller & Heller (2016, § 2.2). Computation of R_n with ties is also $O(n^2)$ (Weihs et al., 2018; Weihs, 2019).

2.2. *Consistency*

In the rest of this section as well as in § 3, we will always assume that the joint distribution function F is continuous, though not necessarily jointly absolutely continuous, with respect to the Lebesgue measure. Accordingly, both $X_1^{(1)}, \ldots, X_n^{(1)}$ and $X_1^{(2)}, \ldots, X_n^{(2)}$ are free of ties with probability 1. To clearly state the following results, we introduce three families of bivariate distributions specified via their joint distribution function F:

$$\begin{aligned} \mathcal{F}^{\mathrm{c}} &= \big\{F : F \text{ is continuous as a bivariate function}\big\}, \\ \mathcal{F}^{\mathrm{ac}} &= \big\{F : F \text{ is absolutely continuous with respect to the Lebesgue measure}\big\}, \\ \mathcal{F}^{\mathrm{DSS}} &= \big\{F \in \mathcal{F}^{\mathrm{c}} : F \text{ has a copula } C(u^{(1)}, u^{(2)}) \text{ that is three- and two-times continuously} \\ &\qquad \text{differentiable with respect to the arguments } u^{(1)} \text{ and } u^{(2)}, \text{ respectively}\big\}. \end{aligned} \tag{6}$$

Recall that the copula of F satisfies $F(x^{(1)}, x^{(2)}) = C\{F_1(x^{(1)}), F_2(x^{(2)})\}$.

We first discuss the large-sample consistency of the correlation coefficients as estimators of the corresponding correlation measures. Convergence in probability is denoted by $\to_{\mathrm{p}}$.

PROPOSITION 2 (CONSISTENCY OF ESTIMATORS). *For any $F \in \mathcal{F}^{\mathrm{c}}$, as $n \to \infty$ we have*

$$\xi_n \to_{\mathrm{p}} \xi, \quad D_n \to_{\mathrm{p}} D, \quad R_n \to_{\mathrm{p}} R, \quad \tau_n^* \to_{\mathrm{p}} \tau^*.$$

If in addition $F \in \mathcal{F}^{\mathrm{DSS}}$ and K, h_1 and h_2 satisfy all the assumptions stated in Definition 2, then also $\xi_n^* \to_{\mathrm{p}} \xi$.

Proof. The claim about ξ_n is Theorem 1.1 in Chatterjee (2021), and the one about ξ_n^* is proved in the Supplementary Material based on a revised version of Theorem 3 in Dette et al. (2013). The remaining claims are immediate from the theory of U-statistics (e.g., Proposition 1 in Weihs et al., 2018, or Theorem 5.4.A in Serfling, 1980). □

Next, we turn to the correlation measures themselves. It is clear that ξ, D and R are always nonnegative, and that the same is true for τ^* when applied to $F \in \mathcal{F}^{\mathrm{c}}$; this follows from Remark 1. The consistency properties for continuous observations can be summarized as follows.

PROPOSITION 3 (CONSISTENCY OF CORRELATION MEASURES). *Each of the correlation measures ξ, R and τ^* is consistent for the entire class $\mathcal{F}^{\mathrm{c}}$; that is, if $F \in \mathcal{F}^{\mathrm{c}}$, then $\xi = 0$, or $R = 0$ or $\tau^* = 0$, if and only if the pair $(X^{(1)}, X^{(2)})$ is independent. Hoeffding's D is consistent for $\mathcal{F}^{\mathrm{ac}}$, but not for $\mathcal{F}^{\mathrm{c}}$.*

Proof. The consistency of ξ is Theorem 2 of Dette et al. (2013) and Theorem 1.1 of Chatterjee (2021). The consistency of R is shown in detail in Theorem 2 of Weihs et al. (2018); see also Blum et al. (1961, p. 490). The consistency of τ^* was established for $\mathcal{F}^{\mathrm{ac}}$ in Theorem 1 of Bergsma & Dassios (2014), and that for $\mathcal{F}^{\mathrm{c}}$ can be shown via Remark 1; cf. Theorem 6.1 of Drton et al. (2020). Finally, the claim about D follows from Theorem 3.1 of Hoeffding (1948) and its generalization in Proposition 3 of Yanagimoto (1970). □

2.3. *Independence tests*

For large samples, computationally efficient independence tests may be implemented using the asymptotic null distributions of the correlation coefficients, which are summarized below.

Proposition 4 (Limiting null distributions). *Suppose that $F \in \mathcal{F}^{\mathrm{c}}$ has $X^{(1)}$ and $X^{(2)}$ independent. As $n \to \infty$, the following properties hold:*

(i) *for Chatterjee's correlation coefficient ξ_n, $n^{1/2}\xi_n \to N(0, 2/5)$ in distribution (Chatterjee, 2021, Theorem* 2.1*);*
(ii) *for Dette–Siburg–Stoimenov's correlation coefficient ξ_n^*, $n^{1/2}\xi_n^* \to 0$ in probability assuming that $F \in \mathcal{F}^{\mathrm{DSS}}$ and that K, h_1 and h_2 satisfy all the assumptions in Definition 2 (revised version of Theorem* 3 *in Dette et al., 2013; see the Supplementary Material);*
(iii) *for $\mu \in \{D, R, \tau^*\}$,*

$$n\mu_n \to \sum_{v_1,v_2=1}^{\infty} \lambda^{\mu}_{v_1,v_2}\left(\xi^2_{v_1,v_2} - 1\right)$$

in distribution, where

$$\lambda^{\mu}_{v_1,v_2} = \begin{cases} 1/(\pi^4 v_1^2 v_2^2), & \mu = D, R, \\ 36/(\pi^4 v_1^2 v_2^2), & \mu = \tau^* \end{cases}$$

for $v_1, v_2 = 1, 2, \ldots$ and $\{\xi_{v_1,v_2}\}$ are independent standard normal random variables (Weihs et al., 2018, Proposition 7*; Drton et al., 2020, Proposition* 3.1*).*

For a given significance level $\alpha \in (0, 1)$, let $z_{1-\alpha/2}$ be the $(1 - \alpha/2)$-quantile of the standard normal distribution. Then the asymptotic test based on Chatterjee's ξ_n is

$$T^{\xi_n}_{\alpha} = I\left\{n^{1/2}|\xi_n| > (2/5)^{1/2} z_{1-\alpha/2}\right\}.$$

The tests based on μ_n with $\mu \in \{D, R, \tau^*\}$ take the form

$$T^{\mu_n}_{\alpha} = I\left(n\,\mu_n > q^{\mu}_{1-\alpha}\right), \quad q^{\mu}_{1-\alpha} = \inf\left[x : \mathrm{pr}\left\{\sum_{v_1,v_2=1}^{\infty} \lambda^{\mu}_{v_1,v_2}\left(\xi^2_{v_1,v_2} - 1\right) \leqslant x\right\} \geqslant 1 - \alpha\right],$$

where $\lambda^{\mu}_{v_1,v_2}$ and ξ_{v_1,v_2} for $v_1, v_2 = 1, \ldots, n, \ldots$ were presented in Proposition 4. Weihs (2019) provides a routine to compute the needed quantiles. It is unclear how to implement the test based on Dette–Siburg–Stoimenov's ξ_n^* without the need for simulation or permutation, as a nondegenerate limiting null distribution is currently unknown.

Given the distribution-free property of ranks for the class $\mathcal{F}^{\mathrm{c}}$, Proposition 4 yields uniform asymptotic validity of the tests just defined. Moreover, Propositions 2 and 3 yield consistency at fixed alternatives. We summarize these facts below.

Proposition 5 (Uniform validity and consistency of tests). *The tests based on the correlation coefficients $\mu_n \in \{\xi_n, D_n, R_n, \tau_n^*\}$ are uniformly valid in the sense that*

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}^{\mathrm{c}}} \mathrm{pr}(T_\alpha^{\mu_n} = 1 \mid H_0) = \alpha.$$

Moreover, these tests are consistent; that is, for fixed $F \in \mathcal{F}^{\mathrm{c}}$ such that $X^{(1)}$ and $X^{(2)}$ are dependent and $\mu_n \in \{\xi_n, R_n, \tau_n^*\}$,

$$\lim_{n\to\infty} \mathrm{pr}(T_\alpha^{\mu_n} = 1 \mid H_1) = 1. \tag{7}$$

The conclusion (7) holds for $\mu_n = D_n$ if it is further assumed that $F \in \mathcal{F}^{\mathrm{ac}}$.

3. Local power analysis

In this section we investigate the local power of the four rank correlation-based tests of H_0 introduced in § 2.3. To this end, we consider two classical and well-used families of alternatives to the null hypothesis of independence: rotation alternatives, or Konijn alternatives (Konijn, 1956), and mixture alternatives, or Farlie-type alternatives (Farlie, 1960, 1961; see also Dhar et al., 2016).

First we consider rotation alternatives. Let $Y^{(1)}$ and $Y^{(2)}$ be two real-valued independent random variables that have mean zero and are absolutely continuous with Lebesgue densities f_1 and f_2, respectively. For $\Delta \in (-1, 1)$, consider

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & \Delta \\ \Delta & 1 \end{pmatrix} \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = A_\Delta \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = A_\Delta Y. \tag{8}$$

For all $\Delta \in (-1, 1)$, the matrix A_Δ is clearly of full rank and invertible. For any $\Delta \in (-1, 1)$, let $f_X(x; \Delta)$ denote the density of $X = A_\Delta Y$. We then make the following assumptions on $Y^{(1)}$ and $Y^{(2)}$.

Assumption 1. The following properties hold:

(i) the distributions of X have a common support for all $\Delta \in (-1, 1)$, so that without loss of generality $\mathcal{X} = \{x : f_X(x; \Delta) > 0\}$ is independent of Δ;
(ii) the density f_k is absolutely continuous with nonconstant logarithmic derivative $\rho_k = f_k'/f_k$ for $k = 1, 2$;
(iii) the Fisher information of X relative to Δ at the point 0, denoted by $\mathcal{I}_X(0)$, is strictly positive, and $E\{(Y^{(k)})^2\} < \infty$ and $E[\{\rho_k(Y^{(k)})\}^2] < \infty$ for $k = 1, 2$.

Remark 4. Assumption 1(ii) and (iii) imply that $E\{\rho_k(Y^{(k)})\} = 0$ and $\mathcal{I}_X(0) < \infty$.

Example 1. Suppose $f_k(z)$, for $k = 1, 2$, is absolutely continuous and positive for all real numbers z. If

$$E\big(Y^{(k)}\big) = 0, \quad E\big\{\big(Y^{(k)}\big)^2\big\} < \infty, \quad E\big[\big\{\rho_k\big(Y^{(k)}\big)\big\}^2\big] < \infty \quad (k = 1, 2), \tag{9}$$

then Assumption 1 holds. As a special case, Assumption 1 holds if $Y^{(1)}$ and $Y^{(2)}$ are centred and follow normal distributions or t-distributions with not necessarily integer-valued degrees of freedom greater than 2.

Second, we consider the following mixture alternatives that were used in Dhar et al. (2016, § 3). Let F_1 and F_2 be fixed univariate distribution functions that are absolutely continuous with Lebesgue density functions f_1 and f_2, respectively. Let $F_0(x^{(1)}, x^{(2)}) = F_1(x^{(1)})F_2(x^{(2)})$ be the product distribution function yielding independence, and let $G \neq F_0$ be a fixed bivariate distribution function which is absolutely continuous and such that $(X^{(1)}, X^{(2)})$ are dependent under G. Let the density functions of F_0 and G, denoted by f_0 and g, respectively, be continuous and have compact supports. Then define the following alternative model for the distribution of $X = (X^{(1)}, X^{(2)})$:

$$F_X = (1 - \Delta)F_0 + \Delta G, \tag{10}$$

with $0 \leqslant \Delta \leqslant 1$.

We make the following additional assumptions on F_0 and G.

Assumption 2. The following properties hold:

(i) the distribution G is absolutely continuous with respect to F_0 and $s(x) = g(x)/f_0(x) - 1$ is continuous;
(ii) the conditional expectation $E\{s(Y) \mid Y^{(1)}\} = 0$ almost surely for $Y = (Y^{(1)}, Y^{(2)}) \sim F_0$;
(iii) the function s is not additively separable, i.e., there do not exist univariate functions h_1 and h_2 such that $s(x) = h_1(x^{(1)}) + h_2(x^{(2)})$;
(iv) the Fisher information $\mathcal{I}_X(0) > 0$.

Remark 5. In this model, $g(x)/f_0(x)$ is continuous and has compact support, which guarantees that $\mathcal{I}_X(0) < \infty$.

Example 2 (*Farlie alternatives*). Let G in (10) be given as

$$G\big(x^{(1)}, x^{(2)}\big) = F_1\big(x^{(1)}\big)F_2\big(x^{(2)}\big)\big[1 + \big\{1 - F_1\big(x^{(1)}\big)\big\}\big\{1 - F_2\big(x^{(2)}\big)\big\}\big].$$

Then Assumption 2 is satisfied (Morgenstern, 1956; Gumbel, 1958; Farlie, 1960). Notice also that $E\{s(Y) \mid Y^{(2)}\} = 0$ almost surely for $Y = (Y^{(1)}, Y^{(2)}) \sim F_0$.

Example 3. Let the density f_2 be symmetric around 0, and consider two univariate functions h_1 and h_2 that are both nonconstant and bounded by 1 in magnitude, with h_2 additionally being an odd function. Let f_1 be a density such that $\int f_1(x^{(1)})h_1(x^{(1)})\, \mathrm{d}x^{(1)} \neq 0$. Then the bivariate density g can be chosen such that $s(x) = h_1(x^{(1)})h_2(x^{(2)})$, and so Assumption 2 holds. For example, we can take $f_1(t) = f_2(t) = 1/2 \times I(-1 \leqslant t \leqslant 1)$, $h_1(t) = |1 - 2\Psi(t)|$ and $h_2(t) = 1 - 2\Psi(t)$, where Ψ denotes the distribution function of the uniform distribution on $[-1, 1]$. In this case, $E\{s(Y) \mid Y^{(2)}\}$ is not almost surely zero for $Y = (Y^{(1)}, Y^{(2)}) \sim F_0$.

For a local power analysis in either of the two alternative families under consideration, we examine the asymptotic power along a respective sequence of alternatives obtained as

$$H_{1,n}(\Delta_0) : \Delta = \Delta_n, \quad \Delta_n = n^{-1/2}\Delta_0, \tag{11}$$

with some constant $\Delta_0 > 0$. We obtain the following results on the discussed tests.

THEOREM 1 (POWER ANALYSIS). *Suppose that the sequences of local alternatives considered are formed such that Assumption 1 or 2 holds when considering a family of rotation or mixture alternatives, respectively. Then concerning any sequence of alternatives given in* (11)*:*

(i) *for either of the two types of alternatives* (A) *and* (B) *and for any fixed constant* $\Delta_0 > 0$*,*

$$\lim_{n\to\infty} \mathrm{pr}\{T_\alpha^{\xi_n} = 1 \mid H_{1,n}(\Delta_0)\} = \alpha; \tag{12}$$

(ii) *for any local alternative family and any number* $\beta > 0$*, there exists a sufficiently large constant* $C_\beta > 0$*, depending only on* β*, such that as long as* $\Delta_0 > C_\beta$*,*

$$\lim_{n\to\infty} \mathrm{pr}\{T_\alpha^{\mu_n} = 1 \mid H_{1,n}(\Delta_0)\} \geqslant 1 - \beta, \tag{13}$$

where $\mu_n \in \{D_n, R_n, \tau_n^*\}$.

In contrast to Theorem 1, Proposition 6 below shows that the power of any size-α test can be arbitrarily close to α when Δ_0 is sufficiently small in the local alternative model $H_{1,\,n}(\Delta_0)$. This result, combined with (12) and (13), entails that the size-α tests based on one of D_n, R_n and τ_n^* are rate-optimal against the local alternatives considered, while the size-α test based on Chatterjee's correlation coefficient, with only trivial power against the local alternative model $H_{1,\,n}(\Delta_0)$ for any fixed Δ_0, is rate-suboptimal.

PROPOSITION 6 (RATE-OPTIMALITY). *Concerning either of the two local alternative families and any sequence of alternatives given in* (11)*, as long as the corresponding Assumption 1 or 2 holds, we have that for any number* $\beta > 0$ *satisfying* $\alpha + \beta < 1$ *there exists a constant* $c_\beta > 0$*, depending only on* β*, such that*

$$\inf_{\bar{T}_\alpha \in \mathcal{T}_\alpha} \mathrm{pr}\{\bar{T}_\alpha = 0 \mid H_{1,n}(c_\beta)\} \geqslant 1 - \alpha - \beta$$

for all sufficiently large n. Here the infimum is taken over all size-α *tests.*

Remark 6. Assumptions 1 and 2 are technical conditions imposed to ensure that (i) the two sequences of alternatives considered are all locally asymptotically normal (van der Vaart, 1998, Ch. 7), i.e., the loglikelihood ratio processes admit a quadratic expansion; (ii) the conditional expectation of the score function given the first margin is almost surely zero. Here the second requirement was invoked to allow use of the conditional multiplier central limit theorem (cf. van der Vaart & Wellner, 1996, § 2.9), which appears to be the key to analysing the power of Chatterjee's correlation coefficient. In addition to their generality, we would like to emphasize that these technical assumptions are indeed satisfied by important models such as Gaussian rotation and Farlie alternatives, which are commonly used to investigate the local power of independence tests.

Remark 7. The linear, step-function, W-shaped, sinusoid, and circular alternatives considered in Chatterjee (2021, § 4.3) can all be viewed as generalized rotation alternatives. The proof techniques used in this paper are therefore directly applicable to these five alternatives by means of a reparameterization. To illustrate this point, consider, for example, the following alternative motivated by Chatterjee (2021, § 4.3):

$$X^{(1)} = Y^{(1)}, \quad X^{(2)} = \Delta g(Y^{(1)}) + Y^{(2)}, \tag{14}$$

where $Y^{(1)}$ and $Y^{(2)}$ are independent and absolutely continuous with respective densities f_1 and f_2. Model (14) and the one used in Chatterjee (2021, § 4.3) are equivalent for rank-based tests as ranks are scale-invariant. Assume then that

(i) the distributions of $X = (X^{(1)}, X^{(2)})$ have a common support for all $\Delta \in (-1, 1)$;
(ii) the density f_2 is absolutely continuous with nonconstant logarithmic derivative $\rho_2 = f_2'/f_2$ such that $0 < E[\{\rho_2(Y^{(2)})\}^2] < \infty$;
(iii) the function g is nonconstant and measurable such that $0 < E[\{g(Y^{(1)})\}^2] < \infty$.

Claims (12) and (13) will then hold for the alternatives (14) in view of arguments similar to those in the proof of Theorem 1 for the rotation alternatives.

Remark 8. Cao & Bickel (2020, § 4.4) performed a local power analysis for Chatterjee's ξ_n under a set of assumptions different from ours. The goal of our local power analysis was to exhibit explicitly the, at times surprising, differences in power of the independence tests given by the four rank correlation coefficients from Definitions 1 and 3–5. We have focused on rotation and mixture alternatives from the literature. However, from the proof techniques in the Supplementary Material, it is evident that (12) and (13) hold for other types of local alternative families. For the former claim, which concerns the lack of power of Chatterjee's ξ_n, this point has been pursued in Cao & Bickel (2020, § 4.4).

4. Rank correlations for discontinuous distributions

In this section, we drop the continuity assumption on F made in § 2 and § 3, and allow ties to exist with nonzero probability. Among the five correlation coefficients, ξ_n^* is no longer an appropriate estimator when F is not continuous. We will discuss the properties of only the other four estimators, ξ_n, D_n, R_n and τ_n^*.

Recall that the computational issue has been addressed in Remark 3. Our first result in this section focuses on approximation consistency of the correlation coefficients ξ_n, D_n, R_n and τ_n^* with respect to their population quantities. To this end, we define the families of distributions to be more general than the ones considered so far:

$$
\begin{aligned}
\mathcal{F} &= \big\{F : F \text{ is a bivariate distribution function}\big\},\\
\mathcal{F}^* &= \big\{F : F_k \text{ is not degenerate, i.e., } F_k(x) \neq I(x \geqslant x_0)\ (k = 1, 2) \text{ for any real number } x_0\big\},\\
\mathcal{F}^{\tau^*} &= \big\{F : F \text{ is discrete, continuous, or a mixture of discrete and jointly absolutely}\\
&\qquad\qquad \text{continuous distribution functions}\big\}.
\end{aligned} \tag{15}
$$

For the estimators ξ_n, D_n, R_n and τ_n^*, we have the following result on consistency.

Proposition 7 (Consistency of estimators). *As $n \to \infty$, the following hold:*

(i) *for $F \in \mathcal{F}^*$, ξ_n converges in probability to ξ (Chatterjee, 2021, Theorem* 1.1*);*
(ii) *for $F \in \mathcal{F}$, μ_n converges in probability to μ for $\mu \in \{D, R, \tau^*\}$ (Weihs et al., 2018, Proposition* 1*; Serfling, 1980, Theorem* 5.4.*A).*

The following proposition is a generalization of Proposition 3.

Proposition 8 (Consistency of correlation measures). *The following are true:*

(i) *for* $F \in \mathcal{F}^*$, $\xi \geqslant 0$ *with equality if and only if the pair is independent (Chatterjee, 2021, Theorem* 1.1*);*
(ii) *for* $F \in \mathcal{F}$ *we have* $D \geqslant 0$*, and for* $F \in \mathcal{F}^{\mathrm{ac}}$ *we have* $D = 0$ *if and only if the pair is independent (Hoeffding, 1948, Theorem* 3.1*; Yanagimoto, 1970, Proposition* 3*);*
(iii) *for* $F \in \mathcal{F}$, $R \geqslant 0$ *with equality if and only if the pair is independent (Blum et al., 1961, p. 490);*
(iv) *for* $F \in \mathcal{F}^{\tau^*}$, $\tau^* \geqslant 0$ *with equality if and only if the variables are independent (Bergsma & Dassios, 2014, Theorem* 1*; Drton et al., 2020, Theorem* 6.1*).*

The asymptotic distribution theory from § 2.3 can also be extended. As the continuity requirement is dropped, the central limit theorem for Chatterjee's ξ_n still holds. However, the asymptotic variance now has a more complicated form and is not necessarily constant across the null hypothesis of independence (Chatterjee, 2021, Theorem 2.2). A similar phenomenon arises for the limiting null distributions of D_n, R_n and τ_n^* when one or two marginals are not continuous; see Nandy et al. (2016, Theorem 4.5 and Corollary 4.1) for further discussion. As a result, permutation analysis, which is unfortunately computationally much more intensive, is typically employed to implement a test outside the realm of continuous distributions.

5. Simulation results

To further examine the power of the tests, we simulate data as a sample consisting of n independent copies of $(X^{(1)}, X^{(2)})$, for which we consider a suite of different specifications based on mixture, rotation and generalized rotation alternatives.

Example 4. For the distribution of $(X^{(1)}, X^{(2)})$ we choose six alternatives. In their specification, $Y^{(1)}$ and $Y^{(2)}$ are always independent random variables and $\Delta = n^{-1/2}\Delta_0$.

(a) The pair $(X^{(1)}, X^{(2)})$ is given by the rotation alternative (8), where $Y^{(1)}$ and $Y^{(2)}$ are both standard Gaussian and $\Delta_0 = 2$. This is an instance of our Example 1.
(b) The pair $(X^{(1)}, X^{(2)})$ is given by the mixture alternative (10), where

$$F_0\big(x^{(1)}, x^{(2)}\big) = \Psi\big(x^{(1)}\big)\Psi\big(x^{(2)}\big),$$
$$G\big(x^{(1)}, x^{(2)}\big) = \Psi\big(x^{(1)}\big)\Psi\big(x^{(2)}\big)\big[1 + \big\{1 - \Psi\big(x^{(1)}\big)\big\}\big\{1 - \Psi\big(x^{(2)}\big)\big\}\big],$$

with $\Psi(\cdot)$ denoting the distribution function of the uniform distribution on $[-1, 1]$, and $\Delta_0 = 10$. This is in accordance with our Example 2.
(c) The pair $(X^{(1)}, X^{(2)})$ is given by the mixture alternative (10), where the density functions of F and G, denoted by f_0 and g, are

$$f_0\big(x^{(1)}, x^{(2)}\big) = \psi\big(x^{(1)}\big)\psi\big(x^{(2)}\big),$$
$$g\big(x^{(1)}, x^{(2)}\big) = \psi\big(x^{(1)}\big)\psi\big(x^{(2)}\big)\big[1 + \big|1 - 2\Psi\big(x^{(1)}\big)\big|\big\{1 - 2\Psi\big(x^{(2)}\big)\big\}\big],$$

with $\psi(t) = 1/2 \times I(-1 \leqslant t \leqslant 1)$, and $\Delta_0 = 20$. This is an instance of our Example 3.
(d) The pair $(X^{(1)}, X^{(2)})$ is given by the generalized rotation alternative (14), where $Y^{(1)}$ is uniformly distributed on $[-1, 1]$, $Y^{(2)}$ is standard Gaussian, g takes values -3, 2, -4 and -3 in the intervals $[-1, -0.5)$, $[-0.5, 0)$, $[0, 0.5)$ and $[0.5, 1]$, respectively, and $\Delta_0 = 3$.

Table 1. *Comparison of the computation time, measured as the total time in seconds of* 1000 *replicates, for all five correlation statistics*

n	ξ_n	ξ_n^*	D_n	R_n	τ_n^*
500	0.157	12.57	0.158	0.263	0.253
1000	0.239	33.75	0.267	0.505	0.468
5000	1.655	401.4	1.823	3.601	3.087
10000	3.089	1152.6	3.315	7.607	7.132

(e) The pair $(X^{(1)}, X^{(2)})$ is given by (14), where $Y^{(1)}$ is uniformly distributed on $[-1, 1]$, $Y^{(2)}$ is standard Gaussian, $g(t) = |t + 0.5|I(t < 0) + |t - 0.5|I(t \geqslant 0)$, and $\Delta_0 = 60$.

(f) The pair $(X^{(1)}, X^{(2)})$ is given by (14), where $Y^{(1)}$ is uniformly distributed on $[-1, 1]$, $Y^{(2)}$ is standard Gaussian, $g(t) = \cos(2\pi t)$, and $\Delta_0 = 12$.

As indicated, the first three simulation settings are taken from Examples 1–3. The latter three are motivated by step-function, W-shaped and sinusoid settings in which Chatterjee's correlation coefficient performs well; see Chatterjee (2021, § 4.3).

Our focus is on comparing the empirical performance of the five tests $T_\alpha^{\xi_n}$, $T_\alpha^{\xi_n^*}$, $T_\alpha^{D_n}$, $T_\alpha^{R_n}$ and $T_\alpha^{\tau_n^*}$. The first four tests are conducted using the asymptotics from Proposition 4. The last test is implemented with bandwidths chosen as $h_1 = h_2 = n^{-3/10}$ following the suggestion in Dette et al. (2013, § 6.1) and using a finite-sample critical value, which we approximate via 1000 Monte Carlo simulations. The nominal significance level is set to 0.05, and the sample size is chosen as $n \in \{500, 1000, 5000, 10\,000\}$. For each of the six settings and four sample sizes, we conduct 1000 simulations.

Before examining the statistical properties, we compare the computation times for calculating the five rank correlation coefficients considered. Table 1 shows times in the rotation setting (a); the results for other settings are essentially the same. The calculations of ξ_n and ξ_n^* are by our own implementation, while those of D_n, R_n and τ_n^* are done using the functions `.calc.hoeffding()`, `.calc.refined()` and `.calc.taustar()`, respectively, from the `R` (R Development Core Team, 2022) package `independence` (Even-Zohar, 2020a). All experiments are conducted on a laptop with a 2.6 GHz Intel Core i5 processor and 8 GB of memory. One can observe the clear computational advantages of ξ_n, D_n, R_n and τ_n^* over Dette et al.'s estimator ξ_n^*. The difference in computation time between Chatterjee's coefficient ξ_n and Hoeffding's D_n is insignificant. Both ξ_n and D_n are slightly faster to compute than Blum–Kiefer–Rosenblatt's R_n and Bergsma–Dassios–Yanagimoto's τ_n^*; computation times differ by a factor less than 2.5.

Table 2 shows the empirical powers of the five tests. The results confirm our earlier theoretical claims about the powers of the different tests in the different models; that is, Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^* outperform Chatterjee's correlation coefficient in all the settings considered. Interestingly, the simulation results suggest that the test based on ξ_n^* may have nontrivial power against certain alternatives; see the results for Example 4(e) and (f) in Table 2.

6. Conclusion

The main new contribution of this work is a local power analysis for continuous distributions that reveals interesting differences in the powers of the tests. The take-away message is that ξ_n is

Table 2. *Empirical powers, based on* 1000 *replicates, of the five competing tests in Example* 4

n	ξ_n	ξ_n^*	D_n	R_n	τ_n^*	ξ_n	ξ_n^*	D_n	R_n	τ_n^*
			Results for Example 4(a)					Results for Example 4(d)		
500	0.103	0.178	0.954	0.955	0.957	0.443	0.122	0.913	0.921	0.919
1000	0.067	0.106	0.956	0.956	0.956	0.285	0.111	0.923	0.928	0.927
5000	0.043	0.078	0.953	0.952	0.952	0.081	0.083	0.936	0.936	0.937
10000	0.045	0.058	0.951	0.952	0.952	0.081	0.052	0.955	0.954	0.955
			Results for Example 4(b)					Results for Example 4(e)		
500	0.087	0.138	0.898	0.896	0.897	0.719	1.000	0.654	0.635	0.643
1000	0.067	0.089	0.900	0.900	0.899	0.486	1.000	0.700	0.682	0.692
5000	0.059	0.082	0.891	0.890	0.891	0.146	1.000	0.735	0.735	0.736
10000	0.052	0.045	0.911	0.914	0.915	0.105	0.997	0.754	0.752	0.752
			Results for Example 4(c)					Results for Example 4(f)		
500	0.088	0.559	0.412	0.404	0.410	0.688	1.000	0.635	0.603	0.611
1000	0.066	0.408	0.390	0.391	0.396	0.459	1.000	0.669	0.655	0.660
5000	0.060	0.327	0.363	0.364	0.364	0.141	1.000	0.717	0.712	0.713
10000	0.048	0.248	0.392	0.395	0.396	0.100	0.994	0.726	0.730	0.728

Table 3. *Properties of the five rank correlation coefficients in Definitions* 1–5*; the bivariate distribution families are defined in* (6) *and* (15)

	μ_n		ξ_n	ξ_n^*	D_n	R_n	τ_n^*
(i)	Computational	$F \in \mathcal{F}^{\mathrm{c}}$	$O(n \log n)$	$O(n^{5/3})$	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$
	efficiency	$F \in \mathcal{F}$	$O(n \log n)$	—	$O(n \log n)$	$O(n^2)$	$O(n^2)$
(ii)	Consistency of correlation measures		$F \in \mathcal{F}^*$	$F \in \mathcal{F}^*$	$F \in \mathcal{F}^{\mathrm{ac}}$	$F \in \mathcal{F}$	$F \in \mathcal{F}^{\tau^*}$
(ii′)	Consistency of independence tests		$F \in \mathcal{F}^{\mathrm{c}}$	$F \in \mathcal{F}^{\mathrm{DSS}}$	$F \in \mathcal{F}^{\mathrm{ac}}$	$F \in \mathcal{F}^{\mathrm{c}}$	$F \in \mathcal{F}^{\mathrm{c}}$
(iii)	Statistical	rotation	rate-suboptimal	—	rate-optimal	rate-optimal	rate-optimal
	efficiency	mixture	rate-suboptimal	—	rate-optimal	rate-optimal	rate-optimal

suboptimal for testing independence, whereas the more classical D_n, R_n and τ_n^* are rate-optimal in the set-up considered. This said, ξ_n and ξ_n^* have very appealing properties that pertain not to independence, but rather to detection of perfect functional dependence. We refer the reader to Dette et al. (2013) and Chatterjee (2021), as well as Cao & Bickel (2020).

We summarize the properties discussed here in Table 3. When referring to independence tests in this table we assume continuous observations, i.e., $F \in \mathcal{F}^{\mathrm{c}}$. Moreover, when discussing ξ_n^*, we assume additionally that the kernel K and bandwidths h_1 and h_2 satisfy all assumptions stated in Definition 2. The table features two rows for computation, where the first pertains to continuous observations free of ties and the second pertains to arbitrary observations. The third row of the table concerns consistency of correlation measures; refer to (6) and (15) for the definitions of table entries. The fourth row concerns consistency of independence tests assuming $F \in \mathcal{F}^{\mathrm{c}}$. Finally, we summarize the rate-optimality and rate-suboptimality of the five independence tests under two local alternatives, rotation and mixture considered in § 3.

Acknowledgement

We thank Holger Dette for pointing out related literature, as well as Sky Cao, Peter Bickel and Bodhisattva Sen for many helpful comments. We thank the editor, the associate editor, and three referees for their very detailed and constructive comments and suggestions, which have greatly improved the quality of the paper. This project was supported by the U.S. National Science Foundation and the European Research Council.

Supplementary material

Supplementary Material available at *Biometrika* online contains detailed proofs.

References

Albert, M., Laurent, B., Marrel, A. & Meynaoui, A. (2021). Adaptive test of independence based on HSIC measures. *arXiv:* 1902.06441v5.

Angus, J. E. (1995). A coupling proof of the asymptotic normality of the permutation oscillation. *Prob. Eng. Inf. Sci.* **9**, 615–21.

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

Berrett, T. B., Kontoyiannis, I. & Samworth, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* **49**, 2457–90.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21**, 593–600.

Blum, J. R., Kiefer, J. & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–98.

Cao, S. & Bickel, P. J. (2020). Correlations with tailored extremal properties. *arXiv:* 2008.10177v2.

Chao, C.-C., Bai, Z. & Liang, W.-Q. (1993). Asymptotic normality for oscillation of permutation. *Prob. Eng. Inf. Sci.* **7**, 227–35.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Am. Statist. Assoc.* **116**, 2009–22.

Deb, N. & Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *arXiv:* 1909.08733v2.

Dette, H., Siburg, K. F. & Stoimenov, P. A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Statist.* **40**, 21–41.

Dhar, S. S., Dassios, A. & Bergsma, W. (2016). A study of the power and robustness of a new test for independence against contiguous alternatives. *Electron. J. Statist.* **10**, 330–51.

Drton, M., Han, F. & Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48**, 3206–27.

Even-Zohar, C. (2020a). *independence: Fast Rank-Based Independence Testing*. R package version 1.0.1, https://CRAN.R-project.org/package=independence.

Even-Zohar, C. (2020b). independence: Fast rank tests. *arXiv:* 2010.09712v2.

Even-Zohar, C. & Leng, C. (2021). Counting small permutation patterns. In *Proc. 2021 ACM-SIAM Symp. Discrete Algorithms (SODA)*. Philadelphia: Society for Industrial and Applied Mathematics, pp. 2288–302.

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307–23.

Farlie, D. J. G. (1961). The asymptotic efficiency of Daniels's generalized correlation coefficients. *J. R. Statist. Soc.* B **23**, 128–42.

Gamboa, F., Klein, T. & Lagnoux, A. (2018). Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA J. Uncert. Quant.* **6**, 522–48.

Gini, C. (1914). *L'ammontare e la Composizione Della Ricchezza Delle Nazioni*, vol. 62 of *Biblioteca di scienze sociali*. Italy: Fratelli Bocca.

Gumbel, E. J. (1958). Distributions à plusieurs variables dont les marges sont données. *C. R. Acad. Sci. Paris* **246**, 2717–19.

Heller, Y. & Heller, R. (2016). Computing the Bergsma Dassios sign-covariance. *arXiv:* 1605.08732v1.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin* **5**, 181–233.

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

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

Kim, I., Balakrishnan, S. & Wasserman, L. (2020). Minimax optimality of permutation tests. *arXiv:* 2003.13208v1.

Konijn, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.* **27**, 300–23.

Maathuis, M., Drton, M., Lauritzen, S. & Wainwright, M., eds. (2019). *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Boca Raton, Florida: CRC Press.

Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsbl. Math. Statist.* **8**, 234–5.
Nandy, P., Weihs, L. & Drton, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Statist.* **10**, 2287–311.
Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge: Cambridge University Press.
R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.
Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.
Shi, H., Drton, M. & Han, F. (2021). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Am. Statist. Assoc.* **117**, 395–410.
Shi, H., Hallin, M., Drton, M. & Han, F. (2020). Rate-optimality of consistent distribution-free tests of independence based on center-outward ranks and signs. *arXiv:* 2007.02186v1.
Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101.
Spearman, C. (1906). 'Footrule' for measuring correlation. *Brit. J. Psychol.* **2**, 89–108.
van der Vaart, A. W. (1998). *Asymptotic Statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.
van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. New York: Springer.
Weihs, L. (2019). *TauStar: Efficient Computation and Testing of the Bergsma-Dassios Sign Covariance*. R package version 1.1.4, https://CRAN.R-project.org/package=TauStar.
Weihs, L., Drton, M. & Leung, D. (2016). Efficient computation of the Bergsma-Dassios sign covariance. *Comp. Statist.* **31**, 315–28.
Weihs, L., Drton, M. & Meinshausen, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105**, 547–62.
Yanagimoto, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.* **22**, 57–63.

[*Received on* 20 *October* 2020. *Editorial decision on* 16 *April* 2021]