The Annals of Statistics
2020, Vol. 48, No. 6, 3206–3227
https://doi.org/10.1214/19-AOS1926

HIGH-DIMENSIONAL CONSISTENT INDEPENDENCE TESTING WITH MAXIMA OF RANK CORRELATIONS

BY MATHIAS DRTON[1], FANG HAN[2,*] AND HONGJIAN SHI[2,†]

[1]*Department of Mathematics, Technical University of Munich, mathias.drton@tum.de*
[2]*Department of Statistics, University of Washington, [*]fanghan@uw.edu; [†]hongshi@uw.edu*

Testing mutual independence for high-dimensional observations is a fundamental statistical challenge. Popular tests based on linear and simple rank correlations are known to be incapable of detecting nonlinear, nonmonotone relationships, calling for methods that can account for such dependences. To address this challenge, we propose a family of tests that are constructed using maxima of pairwise rank correlations that permit consistent assessment of pairwise independence. Built upon a newly developed Cramér-type moderate deviation theorem for degenerate U-statistics, our results cover a variety of rank correlations including Hoeffding's D, Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^*. The proposed tests are distribution-free in the class of multivariate distributions with continuous margins, implementable without the need for permutation, and are shown to be rate-optimal against sparse alternatives under the Gaussian copula model. As a by-product of the study, we reveal an identity between the aforementioned three rank correlation statistics, and hence make a step towards proving a conjecture of Bergsma and Dassios.

1. Introduction. Let $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ be a random vector taking values in $\mathbb{R}^p$ and having all univariate marginal distributions continuous. This paper is concerned with testing the null hypothesis

$$H_0 : X_1, \dots, X_p \quad \text{are mutually independent,} \tag{1.1}$$

based on n independent realizations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ of $\boldsymbol{X}$. Testing H_0 is a core problem in multivariate statistics that has attracted the attention of statisticians for decades; see, for example, the exposition in Anderson (2003), Chapter 9 or Muirhead (1982), Chapter 11. Traditional methods such as the likelihood ratio test, Roy's largest root test (Roy (1957)) and Nagao's L_2-type test (Nagao (1973)) target the case where the dimension p is small and perform poorly when p is comparable to or even larger than n. A line of recent work seeks to address this issue and develops tests that are suitable for modern applications involving data with large dimension p. This high-dimensional regime is in the focus of our work, which develops distribution theory based on asymptotic regimes where $p = p_n$ increases to infinity with n.

Many tests of independence in high dimensions have been proposed recently. For example, Bai et al. (2009) and Jiang and Yang (2013) derived corrected likelihood ratio tests for Gaussian data. Using covariance/correlation statistics such as Pearson's r, Spearman's ρ and Kendall's τ, Bao, Pan and Zhou (2012), Gao et al. (2017), Han, Xu and Zhou (2018) and Bao (2019) proposed revised versions of Roy's largest root test. Schott (2005) and Leung and Drton (2018) derived corrected Nagao's L_2-type tests. Finally, Jiang (2004), Zhou (2007) and Han, Chen and Liu (2017) proposed tests using the magnitude of the largest pairwise correlation statistics. Subsequently, we shall refer to tests of this latter type as maximum-type tests.

Received May 2019; revised November 2019.
MSC2020 subject classifications. 62G10.
Key words and phrases. Degenerate U-statistics, extreme value distribution, independence test, maximum-type test, rank statistics, rate-optimality.

The aforementioned approaches are largely built on linear and simple rank correlations. These, however, are incapable of detecting more complicated nonlinear, nonmonotone dependences as Hoeffding (1948) noted in his seminal paper. Recent work thus proposed the use of consistent rank (Bergsma and Dassios (2014)), kernel-based (Gretton et al. (2008), Pfister et al. (2018)) and distance covariance/correlation statistics (Székely, Rizzo and Bakirov (2007)). However, much less is known about high-dimensional tests of H_0 that use these more involved statistics. Notable exceptions include Leung and Drton (2018) and Yao, Zhang and Shao (2018). There, the authors combined Nagao's L_2-type methods with rank and distance covariance statistics that in a tour de force are shown to weakly converge to a Gaussian limit under the null. In addition, Yao, Zhang and Shao (2018) proved that an infeasible version of their test is rate-optimal against a Gaussian dense alternative (Gaussian distribution with equal correlation), while still little is known about optimality of Leung and Drton's.

In this paper, we derive maximum-type tests that are counterparts of Leung–Drton and Yao–Zhang–Shao L_2-type ones. As noted in Han, Chen and Liu (2017), Leung and Drton (2018) and Yao, Zhang and Shao (2018), maximum-type tests will be more sensitive to strong but sparse dependence. Designed to assess pairwise independence consistently, our tests are formed using statistics based on pairwise rank correlation measures such as Hoeffding's D (Hoeffding (1948)), Blum–Kiefer–Rosenblatt's R (Blum, Kiefer and Rosenblatt (1961)) and Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios (2014), Yanagimoto (1970)). In particular, assuming the pair of random variables X_i and X_j to have a joint distribution that is not only continuous but also absolutely continuous, these measures all satisfy the following three desirable properties summarized in Weihs, Drton and Meinshausen (2018):

I-consistency. If X_i and X_j are independent, the correlation measure is zero.
D-consistency. If X_i and X_j are dependent, the correlation measure is nonzero.
Monotonic invariance. The correlation measure is invariant to monotone transformations.

We remark that invariance under invertible (and not just monotonic) transformations was considered in work on self-equitable measures of dependence (Kinney and Atwal (2014)). This leads to notions of mutual information whose estimates are different from and usually more challenging to handle than the rank correlation measures we consider here; see Berrett and Samworth (2019) and references therein. Indeed, as we shall review in Section 2, the aforementioned correlation measures are naturally estimated via U-statistics, which despite being degenerate have important special properties.

The contributions of our work are threefold. First, we prove that all the maximum-type test statistics proposed in Section 3 have a null distribution that converges to a (nonstandard) Gumbel distribution under high-dimensional asymptotics. This is in contrast to the results in Han, Chen and Liu (2017), where those rank correlation measures that permit consistent assessment of pairwise independence are excluded from the analysis. This exclusion is due to the lack of necessary probability tools like Cramér-type moderate deviation bounds for degenerate U-statistics, which are newly developed in this paper. Additionally, no distributional assumption except for marginal continuity is required for this result, and the parameters for the Gumbel limit can be explicitly given. This allows one to avoid permutation analysis in problems of larger scale. Second, we conduct a power analysis and give explicit conditions on a sparse local alternative under which our proposed tests have power tending to one. Third, we show that the maximum-type tests based on Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^* are all rate-optimal in the class of Gaussian (copula) distributions with sparse and strong dependence as characterized in the power analysis. To our knowledge, this is the first rate-optimality result for a feasible test that permits consistent assessment of pairwise independence. These results are developed in Section 4. The theoretical advantages of our tests are highlighted in simulation studies (Section 5). Lastly, we note that, as an interesting by-product of the study, we give an identity among the above three statistics that helps make a step toward proving Bergsma–Dassios's conjecture about general D-consistency of τ^*. This observation, along with other discussions, is given in Section 6. All proofs and additional simulation results are deferred to the Supplementary Material (Drton, Han and Shi (2020)).

Notation. The sets of real, integer and positive integer numbers are denoted $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{Z}^+$, respectively. The cardinality of a set $\mathcal{A}$ is written $\#\mathcal{A}$. For $m \in \mathbb{Z}^+$, we define $[m] = \{1, 2, \dots, m\}$ and write $\mathcal{P}_m$ for the set of all $m!$ permutations of $[m]$. Let $\boldsymbol{v} = (v_1, \dots, v_p)^\top \in \mathbb{R}^p$, $\mathbf{M} = [\mathbf{M}_{jk}] \in \mathbb{R}^{p\times p}$, and I, J be two subsets of $[p]$. Then $\boldsymbol{v}_I$ is the subvector of $\boldsymbol{v}$ with entries indexed by I, that is, $\boldsymbol{v}_I = (v_{i_1}, v_{i_2}, \dots, v_{i_{\#I}})^\top$ with $i_1 < i_2 < \cdots < i_{\#I}$ and $\{i_1, \dots, i_{\#I}\} = I$. Both $\mathbf{M}_{I,J}$ and $\mathbf{M}[I, J]$ are used to refer to the submatrix of $\mathbf{M}$ with rows indexed by I and columns indexed by J. The matrix $\operatorname{diag}(\mathbf{M}) \in \mathbb{R}^{p\times p}$ is the diagonal matrix whose diagonal is the same as that of $\mathbf{M}$. We write $\mathbf{I}_p$ and $\mathbf{J}_p$ for the identity matrix and all-ones matrix in $\mathbb{R}^{p\times p}$, respectively. For a function $f : \mathcal{X} \to \mathbb{R}$, we define $\|f\|_\infty := \max_{x\in\mathcal{X}} |f(x)|$. The greatest integer less than or equal to $x \in \mathbb{R}$ is denoted by $\lfloor x \rfloor$. The symbol $\mathbb{1}(\cdot)$ is used for indicator functions. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$, $a_n = O(b_n)$, or equivalently $b_n \gtrsim a_n$, if there exists $C > 0$ such that $|a_n| \le C|b_n|$ for any large enough n. We write $a_n \asymp b_n$ if both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold. Write $a_n = o(b_n)$ if for any $c > 0$, $|a_n| \le c|b_n|$ holds for any large enough n. Throughout, c and C refer to positive absolute constants whose values may differ in different parts of the paper.

2. Rank correlations and degenerate U-statistics.

This section introduces the pairwise rank correlations that will later be aggregated in a maximum-type test of the independence hypothesis in (1.1). We present these correlations in a general U-statistic framework. In the sequel, unless otherwise stated, the random vector $\boldsymbol{X}$ is assumed to have continuous margins, that is, its marginal distributions are continuous, though not necessarily absolutely continuous.

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent copies of $\boldsymbol{X}$, with $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,p})^\top$. Let $j \ne k \in [p]$, and let $h : (\mathbb{R}^2)^m \to \mathbb{R}$ be a fixed kernel of order m. The kernel h defines a *U-statistic* of order m:

$$\widehat{U}_{jk} = \binom{n}{m}^{-1} \sum_{1\le i_1<i_2<\cdots<i_m\le n} h\left\{\begin{pmatrix} X_{i_1,j} \\ X_{i_1,k}\end{pmatrix}, \dots, \begin{pmatrix} X_{i_m,j} \\ X_{i_m,k}\end{pmatrix}\right\}. \tag{2.1}$$

For our purposes, h may always be assumed to be *symmetric*, that is, $h(\boldsymbol{z}_1, \dots, \boldsymbol{z}_m) = h(\boldsymbol{z}_{\sigma(1)}, \dots, \boldsymbol{z}_{\sigma(m)})$ for all permutations $\sigma \in \mathcal{P}_m$ and $\boldsymbol{z}_1, \dots, \boldsymbol{z}_m \in \mathbb{R}^2$. Letting $\boldsymbol{z}_i = (z_{i,1}, z_{i,2})^\top$, if both vectors $(z_{1,1}, \dots, z_{m,1})$ and $(z_{1,2}, \dots, z_{m,2})$ are free of ties, that is, have marginal distinct entries, then we have well-defined vectors of ranks $(r_{1,1}, \dots, r_{m,1})$ and $(r_{1,2}, \dots, r_{m,2})$, and we define $\boldsymbol{r}_i = (r_{i,1}, r_{i,2})^\top$ for $1 \le i \le n$. Now a kernel is *rank-based* if

$$h(\boldsymbol{z}_1, \dots, \boldsymbol{z}_m) = h(\boldsymbol{r}_1, \dots, \boldsymbol{r}_m)$$

for all $\boldsymbol{z}_1, \dots, \boldsymbol{z}_m \in \mathbb{R}^2$ with $(z_{1,1}, \dots, z_{m,1})$ and $(z_{1,2}, \dots, z_{m,2})$ free of ties. In this case, we also say that the "correlation" statistic $\widehat{U}_{jk}$ as well as the corresponding "correlation measure" $\mathbb{E}\widehat{U}_{jk}$ is rank-based.

Rank-based statistics have many appealing properties with regard to independence. The following three will be of particular importance for us. Proofs can be found in, for example, Chapter 31 in Kendall and Stuart (1979), Lemma C4 in the supplement of Han, Chen and Liu (2017) and Lemma 2.1 in Leung and Drton (2018). We also note that, in finite samples, the statistics $\{\widehat{U}_{jk}; j < k\}$ are generally not mutually independent.

PROPOSITION 2.1. *Under the null hypothesis in* (1.1) *and assuming continuous margins, we have*:

(i) *The rank statistics* $\{\widehat{U}_{jk}, j \neq k\}$ *are all identically distributed and are distribution-free, that is, the distribution of* $\widehat{U}_{jk}$ *does not depend on the marginal distributions of* $X_1, \dots, X_p$;

(ii) *Fix any* $j \in [p]$, *then the rank statistics* $\{\widehat{U}_{jk}, k \neq j\}$, *are mutually independent*;

(iii) *For any* $j \neq k \in [p]$, *the rank statistic* $\widehat{U}_{jk}$ *is independent of* $\{\widehat{U}_{j'k'}; j', k' \notin \{j,k\}, j' \neq k'\}$.

Our focus will be on those rank-based correlation statistics and the corresponding measures that are induced by the kernel $h(\cdot)$ and are both I- and D-consistent. The kernels of these measures satisfy important additional properties that we will assume in our general treatment. Further concepts concerning U-statistics are needed to state this assumption. For any kernel $h(\cdot)$, any number $\ell \in [m]$, and any measure $\mathbb{P}_{\mathbf{Z}}$, we write

$$h_\ell(z_1 \dots, z_\ell; \mathbb{P}_{\mathbf{Z}}) := \mathbb{E}h(z_1 \dots, z_\ell, \mathbf{Z}_{\ell+1}, \dots, \mathbf{Z}_m)$$

and

$$h^{(\ell)}(z_1, \dots, z_\ell; \mathbb{P}_{\mathbf{Z}}) := h_\ell(z_1, \dots, z_\ell; \mathbb{P}_{\mathbf{Z}}) - \mathbb{E}h - \sum_{k=1}^{\ell-1} \sum_{1 \le i_1 < \dots < i_k \le \ell} h^{(k)}(z_{i_1}, \dots, z_{i_k}; \mathbb{P}_{\mathbf{Z}}), \tag{2.2}$$

where $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ are m independent random vectors with distribution $\mathbb{P}_{\mathbf{Z}}$ and $\mathbb{E}h := \mathbb{E}h(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$. The kernel as well as the corresponding U-statistic is *degenerate* under $\mathbb{P}_{\mathbf{Z}}$ if $h_1(\cdot)$ has variance zero. We use the term *completely degenerate* to indicate that the variances of $h_1(\cdot), \dots, h_{m-1}(\cdot)$ are all zero. Finally, let $\mathbb{P}_0$ be the uniform distribution on $[0,1]$, and write $\mathbb{P}_0 \otimes \mathbb{P}_0$ for its product measure, the uniform distribution on $[0,1]^2$. Note that by Proposition 2.1(i), the study of $\widehat{U}_{jk}$ under independent continuous margins X_j and X_k can be reduced to the case with $(X_j, X_k)^\top \sim \mathbb{P}_0 \otimes \mathbb{P}_0$.

ASSUMPTION 2.1. The kernel h is rank-based, symmetric, and has the following three properties:

(i) h is bounded.

(ii) h is mean-zero and degenerate under independent continuous margins, that is, $\mathbb{E}\{h_1(\mathbf{Z}_1; \mathbb{P}_0 \otimes \mathbb{P}_0)\}^2 = 0$ as $\mathbf{Z}_1 \sim \mathbb{P}_0 \otimes \mathbb{P}_0$.

(iii) $h_2(z_1, z_2; \mathbb{P}_0 \otimes \mathbb{P}_0)$ has uniformly bounded eigenfunctions, that is, it admits the expansion

$$h_2(z_1, z_2; \mathbb{P}_0 \otimes \mathbb{P}_0) = \sum_{v=1}^{\infty} \lambda_v \phi_v(z_1) \phi_v(z_2),$$

where $\{\lambda_v\}$ and $\{\phi_v\}$ are the eigenvalues and eigenfunctions satisfying the integral equation

$$\mathbb{E}h_2(z_1, \mathbf{Z}_2)\phi(\mathbf{Z}_2) = \lambda \phi(z_1) \quad \text{for all } z_1 \in \mathbb{R}^2,$$

with $\mathbf{Z}_2 \sim \mathbb{P}_0 \otimes \mathbb{P}_0$, $\lambda_1 \ge \lambda_2 \ge \dots \ge 0$, $\Lambda := \sum_{v=1}^{\infty} \lambda_v \in (0, \infty)$, and $\sup_v \|\phi_v\|_\infty < \infty$.

The first boundedness property is satisfied for the commonly used rank correlations, including Kendall's τ, Spearman's ρ and many others. The latter two properties are much more specific, but exhibited by the classical rank correlation measures for which consistency properties are known. We discuss the main examples below. Note also that the assumption $\Lambda > 0$ implies $\lambda_1 > 0$, so that $h_2(\cdot)$ is not a constant function.

EXAMPLE 2.1 (Hoeffding's D). From the symmetric kernel,

$$h_D(z_1,\dots,z_5) := \frac{1}{16}\sum_{(i_1,\dots,i_5)\in\mathcal{P}_5}\bigl[\{\mathbb{1}(z_{i_1,1}\le z_{i_5,1})-\mathbb{1}(z_{i_2,1}\le z_{i_5,1})\}\{\mathbb{1}(z_{i_3,1}\le z_{i_5,1})-\mathbb{1}(z_{i_4,1}\le z_{i_5,1})\}\bigr] \times\bigl[\{\mathbb{1}(z_{i_1,2}\le z_{i_5,2})-\mathbb{1}(z_{i_2,2}\le z_{i_5,2})\}\{\mathbb{1}(z_{i_3,2}\le z_{i_5,2})-\mathbb{1}(z_{i_4,2}\le z_{i_5,2})\}\bigr],$$

we recover Hoeffding's D statistic, which is a rank-based U-statistic of order 5 and gives rise to the Hoeffding's D correlation measure $\mathbb{E}h_D$. The kernel $h_D(\cdot)$ satisfies the first two properties in Assumption 2.1 in view of the results in Hoeffding (1948). To verify the last property, we note that under the measure $\mathbb{P}_0\otimes\mathbb{P}_0$, $h_{D,2}(\cdot)$ is known to have eigenvalues

$$\lambda_{i,j;D}=3/(\pi^4 i^2 j^2),\quad i,j\in\mathbb{Z}^+;$$

see, for example, Proposition 7 in Weihs, Drton and Meinshausen (2018) or Theorem 4.4 in Nandy, Weihs and Drton (2016). The corresponding eigenfunctions are

$$\phi_{i,j;D}\{(z_{1,1},z_{1,2})^\top\}=2\cos(\pi i z_{1,1})\cos(\pi j z_{1,2}),\quad i,j\in\mathbb{Z}^+.$$

The eigenvalues are positive and sum to $\Lambda_D:=\sum_{i,j}\lambda_{i,j;D}=1/12$, and $\sup_{i,j}\|\phi_{i,j;D}\|_\infty\le 2$. For any pair of random variables, the correlation measure $\mathbb{E}h_D\ge 0$ (Hoeffding (1948), p. 547). Furthermore, it has been proven that, once the pair is absolutely continuous in $\mathbb{R}^2$, the correlation measure $\mathbb{E}h_D=0$ if and only if the pair is independent (Hoeffding (1948), Yanagimoto (1970)). This property, however, generally does not hold for discrete data or data generated from a bivariate distribution that is continuous but not absolutely continuous; see Remark 1 in Yanagimoto (1970) for a counterexample.

EXAMPLE 2.2 (Blum–Kiefer–Rosenblatt's R). The symmetric kernel

$$h_R(z_1,\dots,z_6) := \frac{1}{32}\sum_{(i_1,\dots,i_6)\in\mathcal{P}_6}\bigl[\{\mathbb{1}(z_{i_1,1}\le z_{i_5,1})-\mathbb{1}(z_{i_2,1}\le z_{i_5,1})\}\{\mathbb{1}(z_{i_3,1}\le z_{i_5,1})-\mathbb{1}(z_{i_4,1}\le z_{i_5,1})\}\bigr] \times\bigl[\{\mathbb{1}(z_{i_1,2}\le z_{i_6,2})-\mathbb{1}(z_{i_2,2}\le z_{i_6,2})\}\{\mathbb{1}(z_{i_3,2}\le z_{i_6,2})-\mathbb{1}(z_{i_4,2}\le z_{i_6,2})\}\bigr]$$

yields Blum–Kiefer–Rosenblatt's R statistic (Blum, Kiefer and Rosenblatt (1961)), which is a rank-based U-statistic of order 6. One can verify the three properties in Assumption 2.1 similarly to Hoeffding's D by using that $h_{R,2}=2h_{D,2}$. In addition, for any pair of random variables, the correlation measure $\mathbb{E}h_R\ge 0$ with equality if and only if the pair is independent, and no continuity assumption is needed at all; cf. page 490 of Blum, Kiefer and Rosenblatt (1961).

EXAMPLE 2.3 (Bergsma–Dassios–Yanagimoto's τ^*). Bergsma and Dassios (2014) introduced a rank correlation statistic as a U-statistic of order 4 with the symmetric kernel

$$\begin{aligned} h_{\tau^*}(z_1,\dots,z_4) := \frac{1}{16}\sum_{(i_1,\dots,i_4)\in\mathcal{P}_4}&\{\mathbb{1}(z_{i_1,1},z_{i_3,1}<z_{i_2,1},z_{i_4,1})+\mathbb{1}(z_{i_2,1},z_{i_4,1}<z_{i_1,1},z_{i_3,1})\\ &-\mathbb{1}(z_{i_1,1},z_{i_4,1}<z_{i_2,1},z_{i_3,1})-\mathbb{1}(z_{i_2,1},z_{i_3,1}<z_{i_1,1},z_{i_4,1})\}\\ &\times\{\mathbb{1}(z_{i_1,2},z_{i_3,2}<z_{i_2,2},z_{i_4,2})+\mathbb{1}(z_{i_2,2},z_{i_4,2}<z_{i_1,2},z_{i_3,2})\\ &-\mathbb{1}(z_{i_1,2},z_{i_4,2}<z_{i_2,2},z_{i_3,2})-\mathbb{1}(z_{i_2,2},z_{i_3,2}<z_{i_1,2},z_{i_4,2})\}. \end{aligned}$$

Here, $\mathbb{1}(y_1, y_2 < y_3, y_4) := \mathbb{1}(y_1 < y_3)\mathbb{1}(y_1 < y_4)\mathbb{1}(y_2 < y_3)\mathbb{1}(y_2 < y_4)$. It holds that $h_{\tau^*,2} = 3h_{D,2}$ and all properties in Assumption 2.1 also hold for $h_{\tau^*}(\cdot)$. Theorem 1 in Bergsma and Dassios (2014) shows that for a pair of random variables whose distribution is discrete, absolutely continuous or a mixture of both, the correlation measure $\mathbb{E}h_{\tau^*} \geq 0$ where equality holds if and only if the variables are independent. It has been conjectured that this fact is true for any distribution on $\mathbb{R}^2$. In Section 6.2 of this paper, we make new progress along this track. This progress is based on early but apparently little known results of Yanagimoto (1970) that prompted us to add his name in reference to τ^*.

3. Maximum-type tests of mutual independence. We now turn to tests of the mutual independence hypothesis H_0 in (1.1). As in Han, Chen and Liu (2017), we propose maximum-type tests. However, in contrast to Han, Chen and Liu (2017), we suggest the use of consistent and rank-based correlations with the practical choices being the ones from Examples 2.1–2.3. As these measures are all nonnegative, it is appropriate to consider a one-sided test in which we aggregate pairwise U-statistics $\widehat{U}_{jk}$ in (2.1) into the test statistic

$$\widehat{M}_n := (n-1)\max_{j<k} \widehat{U}_{jk}.$$

We then reject H_0 if $\widehat{M}_n$ is larger than a certain threshold. Note that we tacitly assumed $\widehat{U}_{jk} = \widehat{U}_{kj}$ when maximizing over $j < k$; this symmetry holds for any reasonable correlation statistic. We emphasize once more that, since the statistic is constructed based on pairs $\{X_{i,j}, X_{i,k}\}_{i\in[n]}$, the proposed tests are designed to assess pairwise independence consistently.

By Proposition 2.1(i), the statistic $\widehat{M}_n$ is distribution-free in the class of multivariate distributions with continuous margins. An exact critical value for rejection of H_0 could thus be approximated by Monte Carlo simulation. However, as we will show, extreme-value theory yields asymptotic critical values that avoid any extra computation all the while giving good finite-sample control of the test's size. When presenting this theory, we write $X \overset{\mathsf{d}}{=} Y$ if two random variables X and Y have the same distribution, and we use $\xrightarrow{\mathsf{d}}$ to denote "weak convergence."

If, under H_0, the studied statistic $(n-1)\widehat{U}_{jk}$ weakly converged to a chi-square distribution with one degree of freedom, as in Theorems 1 and 2 of Han, Chen and Liu (2017), then extreme-value theory combined with Proposition 2.1 would imply that a suitably standardized version of $\widehat{M}_n$ would weakly converge to a type-I Gumbel distribution with distribution function $\exp\{-(8\pi)^{-1/2}\exp(-y/2)\}$. However, the degeneracy stated in Assumption 2.1(ii) rules out this possibility. Classical theory yields that instead of a single chi-square variable, we encounter convergence to much more involved infinite weighted series (Serfling (1980), Chapter 5.5.2).

PROPOSITION 3.1. *Let* $\boldsymbol{X}$ *have continuous margins, and let* $j \neq k$. *If* $h(\cdot)$ *satisfies Assumption* 2.1, *then under* H_0,

$$\binom{m}{2}^{-1}(n-1)\widehat{U}_{jk} \xrightarrow{\mathsf{d}} \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1),$$

where $\{\xi_v, v = 1, 2, \dots\}$ *are i.i.d. standard Gaussian random variables.*

Note that the weak convergence result for degenerate U-statistics in Proposition 3.1 holds under much weaker conditions than Assumption 2.1; see the main theorem in Serfling (1980), Chapter 5.5.2, for detailed conditions. Our intuition for the asymptotic forms of the maxima now comes from the following fact, though the analysis of $\max_{j<k} \widehat{U}_{jk}$ requires more refined techniques since $\{\widehat{U}_{jk}; j \leq k\}$ are in general not mutually independent.

PROPOSITION 3.2. *Let $Y_1, \dots, Y_d$ be $d = p(p-1)/2$ independent copies of $\zeta \stackrel{\mathsf{d}}{=} \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1)$. Then, as $p \to \infty$,*

$$\max_{j\in[d]} \frac{Y_j}{\lambda_1} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} \xrightarrow{\mathsf{d}} G.$$

Here, G follows a Gumbel distribution with distribution function

$$\exp\left\{-\frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\left(-\frac{y}{2}\right)\right\},$$

where μ_1 is the multiplicity of the largest eigenvalue λ_1 in the sequence $\{\lambda_1, \lambda_2, \dots\}$, $\kappa := \prod_{v=\mu_1+1}^{\infty}(1 - \lambda_v/\lambda_1)^{-1/2}$, and $\Gamma(z) := \int_0^\infty x^{z-1}e^{-x}\,dx$ is the gamma function.

Obviously, when setting $\lambda_1 = 1, \lambda_2 = \lambda_3 = \cdots = 0$ in Proposition 3.2, we recover the Gumbel distribution derived by Han, Chen and Liu (2017). Based on Propositions 3.1 and 3.2, for any prespecified significance level $\alpha \in (0, 1)$, our proposed test is

$$\mathsf{T}_\alpha := \mathbb{1}\left\{\frac{n-1}{\lambda_1\binom{m}{2}}\max_{j<k}\widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} > Q_\alpha\right\}, \tag{3.1}$$

where

$$Q_\alpha := \log\frac{2^{\mu_1-4}\kappa^2}{\{\Gamma(\mu_1/2)\}^2} - 2\log\log(1-\alpha)^{-1}$$

is the $1-\alpha$ quantile of the Gumbel distribution of distribution function $\exp\{-2^{\mu_1/2-2}\kappa/\Gamma(\mu_1/2)\cdot\exp(-y/2)\}$. However, note that so far the test results merely from heuristic arguments. Theoretical justifications regarding the test's size and power under the high-dimensional regime will be given in Section 4.

EXAMPLE 3.1 ("Extreme D"). Hoeffding's D statistic introduced in Example 2.1 is

$$\widehat{D}_{jk} := \binom{n}{5}^{-1}\sum_{i_1<\cdots<i_5} h_D(\boldsymbol{X}_{i_1,\{j,k\}}, \dots, \boldsymbol{X}_{i_5,\{j,k\}}).$$

According to (3.1), the corresponding test is

$$\mathsf{T}_{D,\alpha} := \mathbb{1}\left\{\frac{\pi^4(n-1)}{30}\max_{j<k}\widehat{D}_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{D,\alpha}\right\},$$

where $Q_{D,\alpha} := \log\{\kappa_D^2/(8\pi)\} - 2\log\log(1-\alpha)^{-1}$ and

$$\kappa_D := \left\{2\prod_{n=2}^{\infty}\frac{\pi/n}{\sin(\pi/n)}\right\}^{1/2} \approx 2.467.$$

EXAMPLE 3.2 ("Extreme R"). Blum–Kiefer–Rosenblatt's R statistic from Example 2.2 is

$$\widehat{R}_{jk} := \binom{n}{6}^{-1}\sum_{i_1<\cdots<i_6} h_R(\boldsymbol{X}_{i_1,\{j,k\}}, \dots, \boldsymbol{X}_{i_6,\{j,k\}}).$$

According to (3.1), the corresponding test is

$$\mathsf{T}_{R,\alpha} := \mathbb{1}\left\{\frac{\pi^4(n-1)}{90}\max_{j<k}\widehat{R}_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{R,\alpha}\right\},$$

where $Q_{R,\alpha} := Q_{D,\alpha}$.

EXAMPLE 3.3 ("Extreme τ^*"). Bergsma–Dassios–Yanagimoto's τ^* statistic from Example 2.3 is

$$\widehat{\tau}^*_{jk} := \binom{n}{4}^{-1} \sum_{i_1<\cdots<i_4} h_{\tau^*}(\boldsymbol{X}_{i_1,\{j,k\}}, \ldots, \boldsymbol{X}_{i_4,\{j,k\}}).$$

According to (3.1), it yields the test

$$\mathsf{T}_{\tau^*,\alpha} := \mathbb{1}\left\{\frac{\pi^4(n-1)}{54} \max_{j<k} \widehat{\tau}^*_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{\tau^*,\alpha}\right\},$$

where $Q_{\tau^*,\alpha} := Q_{D,\alpha}$.

Note that, by the definitions of the kernels and the identity (6.1) that will be introduced in Section 6.2, as long as there is no tie in the data, for any $j, k \in [p]$,

$$\widehat{D}_{jj} = \widehat{R}_{jj} = \widehat{\tau}^*_{jj} = 1 \quad \text{and} \quad 3\widehat{D}_{jk} + 2\widehat{R}_{jk} = 5\widehat{\tau}^*_{jk}. \tag{3.2}$$

REMARK 3.1. In applying the above tests, we have intrinsically assumed that there are no ties among the entries $X_{1,j}, \ldots, X_{n,j}$ for each $j \in [p]$. This is based on the assumption that $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ has continuous margins. In practice, however, data in finite accuracy might feature ties or may indeed be drawn from a distribution that is not of a continuous margin. In such cases, conducting the above tests on the original data may distort the size. To fix this, as was discussed in Remark 2.1 in Heller et al. (2016), one may break the ties randomly so that the above tests remain distribution-free. Also see Chapter 8 in Hollander, Wolfe and Chicken (2014) for more discussions on how to break ties for rank-based tests.

4. Theoretical analysis. This section provides theoretical justifications of the tests proposed in Section 3. The section is split into two parts. The first part rigorously justifies the proposed asymptotic critical values. The second part gives a power analysis and shows optimality properties.

4.1. *Size control.* In this section, we derive the limiting distribution of the statistic $\widehat{M}_n$ under H_0. The below Cramér-type moderate deviation theorem for degenerate U-statistics under a general probability measure is the foundation of our theory. There has been a large literature on deriving the moderate deviation theorem for nondegenerate U-statistics (see, e.g., Shao and Zhou (2016) for some recent developments) as well as Berry–Esseen-type bounds for degenerate U-statistics (see Zhou and Götze (1997) and Götze and Zaitsev (2014) among many). However, to our knowledge, the literature does not provide a comparable moderate deviation theorem for degenerate U-statistics.

THEOREM 4.1 (Cramér-type moderate deviation for degenerate U-statistics). *Let $Z_1, \ldots, Z_n$ be (not necessarily continuous) i.i.d. random variables with distribution $\mathbb{P}_Z$. Consider the U-statistic*

$$\widehat{U}_n = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} h(Z_{i_1}, \ldots, Z_{i_m}),$$

where the kernel $h(\cdot)$ is symmetric and such that (i) $\|h\|_\infty < \infty$, (ii) $h_1(Z_1; \mathbb{P}_Z) = 0$ almost surely, and (iii) $h_2(z_1, z_2; \mathbb{P}_Z)$ admits the eigenfunction expansion,

$$h_2(z_1, z_2; \mathbb{P}_Z) = \sum_{v=1}^{\infty} \lambda_v \phi_v(z_1)\phi_v(z_2),$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\Lambda := \sum_{v=1}^{\infty} \lambda_v \in (0, \infty)$, *and* $\sup_v \|\phi_v\|_\infty < \infty$. *We then have, for any sequence of positive scalars* $e_n \to 0$,

$$\lim_{n\to\infty} \sup_{x_n \in [-\Lambda, e_n n^\theta]} \left| \frac{\mathbb{P}\{\binom{m}{2}^{-1}(n-1)\widehat{U}_n > x_n\}}{\mathbb{P}\{\sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1) > x_n\}} - 1 \right| = 0,$$

where $\{\xi_v, v = 1, 2, \ldots\}$ *are i.i.d. standard Gaussian, and* θ *is any absolute constant such that*

$$\theta < \sup\left\{q \in [0, 1/3) : \sum_{v > \lfloor n^{(1-3q)/5} \rfloor} \lambda_v = O(n^{-q})\right\} \tag{4.1}$$

if infinitely many of eigenvalues λ_v *are nonzero, and* $\theta = 1/3$ *otherwise.*

In Theorem 4.1, when there are only finitely many nonzero eigenvalues, the range $o(n^{1/3})$ is the standard one for Cramér-type moderate deviation. When there are infinitely many nonzero eigenvalues, it is still unclear if the range $o(n^\theta)$ is the best possible one. It is certainly an interesting question to investigate the optimal range for degenerate U-statistics in the future. With the aid of Theorem 4.1 and combining it with Proposition 3.2, we can now show that, under H_0, even if p is exponentially larger than the sample size n, our maximum-type test statistic still weakly converges to the Gumbel distribution specified in Proposition 3.2. Hence, the proposed test T_α in (3.1) can effectively control the size.

THEOREM 4.2 (Limiting null distribution). *Assume* $X_1, \ldots, X_p$ *are continuous and the independence hypothesis* H_0 *holds. Let* $\widehat{U}_{jk}$, $j < k$, *have a common kernel* h *that satisfies Assumption* 2.1. *Define the parameter* θ *as in* (4.1). *Then if* $p = p_n$ *goes to infinity with* n *such that* $\log p = o(n^\theta)$, *it holds for any absolute constant* $y \in \mathbb{R}$ *that*

$$\mathbb{P}\left\{\frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} \leq y\right\}$$
$$= \exp\left\{-\frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)} \exp\left(-\frac{y}{2}\right)\right\} + o(1).$$

Consequently,

$$\mathbb{P}_{H_0}(\mathsf{T}_\alpha = 1) = \alpha + o(1),$$

where $\mathbb{P}_{H_0}$ *represents the probability under the null hypothesis* H_0.

Note that the proof of Theorem 4.2 uses the Chen–Stein method, via Theorem 1 of Arratia, Goldstein and Gordon (1989), which is able to handle our case where the random variables are not mutually independent. We emphasize that our theory holds without any distributional assumption on $\boldsymbol{X}$ except for marginal continuity. This property of being distribution-free in the class of multivariate distributions with continuous margins is essentially shared by all rank-based correlation measures, but is clearly not satisfied by other measures like linear or distance covariance as was illustrated, for example, by Jiang (2004) and Yao, Zhang and Shao (2018).

As a simple consequence of Theorem 4.2, the following corollary shows that the tests in Examples 3.1–3.3 have asymptotically correct sizes, with θ being explicitly calculated.

COROLLARY 4.1. *Let* $X_1, \ldots, X_p$ *be continuous. Let* p *go to infinity with* n *in such a way that* $\log p = o(n^{1/8-\delta})$ *for some arbitrarily small prespecified constant* $\delta > 0$. *Then*

$$\mathbb{P}_{H_0}(\mathsf{T}_{D,\alpha} = 1) = \alpha + o(1), \qquad \mathbb{P}_{H_0}(\mathsf{T}_{R,\alpha} = 1) = \alpha + o(1) \quad \textit{and}$$
$$\mathbb{P}_{H_0}(\mathsf{T}_{\tau^*,\alpha} = 1) = \alpha + o(1).$$

4.2. *Power analysis and rate-optimality.* We now investigate the power of the proposed tests from an asymptotic minimax perspective. The key ingredient is the choice of a suitable distribution family as an alternative to the null hypothesis in (1.1). Recall the definition of $h^{(1)}(\cdot)$ in (2.2). For any kernel function $h(\cdot)$ and constants $\gamma > 0$ and $q \in \mathbb{Z}^+$, define a general q-dimensional (not necessarily continuous) distribution family as follows:

$$\mathcal{D}(\gamma, q; h) := \{\mathcal{L}(\boldsymbol{X}) : \boldsymbol{X} \in \mathbb{R}^q, \mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_{jk})\} \le \gamma \mathbb{E}_{jk} h \text{ for all } j \ne k \in [q]\},$$

where $\mathcal{L}(\boldsymbol{X})$ is the distribution (law) of $\boldsymbol{X}$, and $\mathbb{P}_{jk}$, $\mathbb{E}_{jk}(\cdot)$, and $\mathrm{Var}_{jk}(\cdot)$ stand for the probability measure, expectation and variance operated on the bivariate distribution of $(X_j, X_k)^\top$, respectively.

The family $\mathcal{D}(\gamma, q; h)$ intrinsically characterizes the slope of the function $\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_{jk})\}$ with regard to the dependence between X_j and X_k, characterized by the "correlation measure" $\mathbb{E}_{jk}h$. Intuitively, consider $\mathbb{E}_{jk}h$ as a rank correlation measure of dependence between X_j and X_k. When X_j is independent of X_k, we have that

$$\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_j \otimes \mathbb{P}_k)\} = 0 = \mathbb{E}_{jk} h$$

as long as Assumption 2.1 holds for $h(\cdot)$. Therefore, heuristically, as the dependence between X_j and X_k increases, it is possible that the variance $\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_{jk})\}$ will deviate from 0 at the same or a slower rate compared to $\mathbb{E}_{jk}h$. Note that both parameters are nonnegative. The next lemma firms up this intuition by establishing that the Gaussian family belongs to $\mathcal{D}(\gamma, q; h)$ for all the kernels $h(\cdot)$ considered in Examples 2.1 to 2.3, provided γ is large enough.

LEMMA 4.1. *There exists an absolute constant $\gamma > 0$ such that for all $q \in \mathbb{Z}^+$, any q-dimensional Gaussian distribution is in $\mathcal{D}(\gamma, q; h_D)$, $\mathcal{D}(\gamma, q; h_R)$, and $\mathcal{D}(\gamma, q; h_{\tau^*})$.*

Next, we introduce a class of matrices indexed by a positive constant C as

$$\mathcal{U}_p(C) := \Big\{\mathbf{M} \in \mathbb{R}^{p \times p} : \max_{j<k}\{M_{jk}\} \ge C(\log p/n)\Big\}.$$

Such matrices will define a "sparse local alternative" as considered also in Section 4.1 in Han, Chen and Liu (2017). Note, however, that in our case the scale is at the order of $\log p/n$ as opposed to $(\log p/n)^{1/2}$ in Han, Chen and Liu (2017). This is due to our statistics being degenerate under independence. Hence, the variance of $h^{(1)}(\cdot)$ is zero under the null, while nonzero for these statistics investigated in Han, Chen and Liu (2017). It should also be noted that these two classes cannot be directly compared; intuitively the consistent measures are defined on a squared scale when contrasted to the nonconsistent measures. As will be shown later, in the example of the Gaussian case, both classes correspond to a condition on the Pearson correlation obeying the rate $(\log p/n)^{1/2}$.

The following theorem now describes "local alternatives" under which the power of our general test T_α tends to one as both n and p go to infinity.

THEOREM 4.3 (Power analysis, general). *Given any $\gamma > 0$ and a kernel $h(\cdot)$ satisfying Assumption 2.1, there exists some sufficiently large C_γ depending on γ such that*

$$\liminf_{n,p\to\infty} \inf_{\mathbf{U}\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{U}}(\mathsf{T}_\alpha = 1) = 1,$$

where, for each specified (n, p), the infimum is taken over all distributions in $\mathcal{D}(\gamma, p; h)$ that have the matrix of population dependence coefficients $\mathbf{U} = [U_{jk}]$ in $\mathcal{U}_p(C_\gamma)$. Here, $U_{jk} := \mathbb{E}\widehat{U}_{jk}$.

The proof of Theorem 4.3 only uses the Hoeffding decomposition for U-statistics, Bernstein's inequality for the sample mean part and Arcones and Giné's inequality for the degenerate U-statistics parts (Arcones and Giné (1993)). Consequently, we do not have to assume any continuity of $\boldsymbol{X}$. The theorem immediately yields the following corollary, characterizing the local alternatives under which the three rank-based tests from Examples 3.1–3.3 have power tending to 1.

COROLLARY 4.2 (Power analysis, examples). *Given any $\gamma > 0$, we have, for some sufficiently large C_γ depending on γ,*

$$\liminf_{n,p\to\infty} \inf_{\mathbf{D}\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{D}}(\mathsf{T}_{D,\alpha}=1)=1, \qquad \liminf_{n,p\to\infty} \inf_{\mathbf{R}\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{R}}(\mathsf{T}_{R,\alpha}=1)=1,$$

$$\liminf_{n,p\to\infty} \inf_{\mathbf{T}^*\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{T}^*}(\mathsf{T}_{\tau^*,\alpha}=1)=1,$$

where, for each specified (n, p), the infima are taken over all distributions in $\mathcal{D}(\gamma, p; h_D)$, $\mathcal{D}(\gamma, p; h_R)$, and $\mathcal{D}(\gamma, p; h_{\tau^*})$ with population dependence coefficient matrices $\mathbf{D} = [D_{jk}]$, $\mathbf{R} = [R_{jk}]$, and $\mathbf{T}^* = [\tau^*_{jk}]$ for $D_{jk} := \mathbb{E}\widehat{D}_{jk}$, $R_{jk} := \mathbb{E}\widehat{R}_{jk}$ and $\tau^*_{jk} := \mathbb{E}\widehat{\tau}^*_{jk}$, respectively.

We now turn to optimality of the proposed tests. There have been long debates on the power of consistent rank-based tests compared to those based on linear and simple rank correlation measures. As a matter of fact, Blum, Kiefer and Rosenblatt (1961) have given interesting comments on this topic, stating that the required sample size for the bivariate independence test based on $h_R(\cdot)$ is of the same order as that in common parametric cases, hinting that even under a particular parametric model these nonparametric consistent tests of independence can be as rate-efficient as tests that specifically target the considered model. Leung and Drton (2018) and Han, Chen and Liu (2017), among many others, derived rate-optimality results for rank-based tests. However, their results do not cover those that permit consistent assessment of pairwise independence. Recently, Yao, Zhang and Shao (2018) made a first step toward a minimax optimality result for consistent tests of independence. Their result shows an infeasible version of a test based on distance covariance to be rate-optimal against a Gaussian dense alternative. However, it remained an open question if there exists a feasible (consistent) test of mutual independence in high dimensions that is rate-optimal against certain alternatives. Below we are able to give an affirmative answer.

We shall focus on the proposed tests in Examples 3.1–3.3 and show their rate-optimality in the Gaussian model. To this end, we define a new alternative class of matrices

$$\mathcal{V}(C) := \left\{\mathbf{M} \in \mathbb{R}^{p\times p} : \mathbf{M} \succeq 0, \operatorname{diag}(\mathbf{M}) = \mathbf{I}_p, \mathbf{M} = \mathbf{M}^\top, \max_{j\neq k} |M_{jk}| \geq C\sqrt{\frac{\log p}{n}}\right\},$$

where $\mathbf{M} \succeq 0$ denotes positive semidefiniteness. We then have the following theorem as a consequence of Corollary 4.2. It concerns the proposed tests' power under a Gaussian model with some nonzero pairwise correlations but for which these are decaying to zero as the sample size increases, and is immediate from the fact that, as $(X_j, X_k)^\top$ is bivariately normal with correlation ρ_{jk}, we have

$$D_{jk}, R_{jk}, \tau^*_{jk} \asymp \rho^2_{jk} \quad \text{as } \rho_{jk} \to 0.$$

Since the test statistics are all rank-based, and thus invariant to monotone marginal transformations, extension of the following result to the corresponding Gaussian copula family with continuous margins is straightforward.

THEOREM 4.4 (Power analysis, Gaussian). *For a sufficiently large absolute constant* $C_0 > 0$, *we have, as long as* $n, p \to \infty$,

$$\inf_{\mathbf{\Sigma} \in \mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{D,\alpha} = 1) = 1 - o(1), \qquad \inf_{\mathbf{\Sigma} \in \mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{R,\alpha} = 1) = 1 - o(1) \quad \text{and}$$

$$\inf_{\mathbf{\Sigma} \in \mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{\tau^*,\alpha} = 1) = 1 - o(1),$$

where infima are over centered Gaussian distributions with (*Pearson*) *covariance matrix* $\mathbf{\Sigma} = [\Sigma_{jk}]$.

The proof of Theorem 4.4 is given in the Supplementary Material. It relies on Lemma 4.1 and the fact that $D_{jk}, R_{jk}, \tau^*_{jk} \asymp \Sigma^2_{jk}$ as $\Sigma_{jk} \to 0$. Combined with the following result from Han, Chen and Liu (2017), Theorem 4.4 yields minimax rate-optimality of the tests in Examples 3.1–3.3 against the sparse Gaussian alternative.

THEOREM 4.5 (Rate optimality, Theorem 5 in Han, Chen and Liu, 2017). *There exists an absolute constant* $c_0 > 0$ *such that for any number* $\beta > 0$ *satisfying* $\alpha + \beta < 1$, *in any asymptotic regime with* $p \to \infty$ *as* $n \to \infty$ *but* $\log p/n = o(1)$, *it holds for all sufficiently large* n *and* p *that*

$$\inf_{\overline{\mathsf{T}}_\alpha \in \mathcal{T}_\alpha} \sup_{\mathbf{\Sigma} \in \mathcal{V}(c_0)} \mathbb{P}_{\mathbf{\Sigma}}(\overline{\mathsf{T}}_\alpha = 0) \geq 1 - \alpha - \beta.$$

Here, the infimum is taken over all size-α *tests, and the supremum is taken over all centered Gaussian distributions with* (*Pearson*) *covariance matrix* $\mathbf{\Sigma}$.

5. Simulation studies. In this section, we compare the finite-sample performance of the three tests (Extreme D, Extreme R and Extreme τ^*) from Section 3 to eight existing tests proposed in the literature via Monte Carlo simulations. The first eight tests are rank-based and hence distribution-free in the class of multivariate distributions with continuous margins, while the other three tests are distribution-dependent:

DHS$_D$: the maximum-type test in Example 3.1;
DHS$_R$: the maximum-type test in Example 3.2;
DHS$_{\tau^*}$: the maximum-type test in Example 3.3;
LD$_\tau$: the L_2-type test based on Kendall's τ (Leung and Drton (2018));
LD$_\rho$: the L_2-type test based on Spearman's ρ (Leung and Drton (2018));
LD$_{\tau^*}$: the L_2-type test based on Bergsma–Dassios–Yanagimoto's τ^* (Leung and Drton (2018));
HCL$_\tau$: the maximum-type test based on Kendall's τ (Han, Chen and Liu (2017));
HCL$_\rho$: the maximum-type test based on Spearman's ρ (Han, Chen and Liu (2017));
YZS: the L_2-type test based on the distance covariance statistic (Yao, Zhang and Shao (2018));
SC: the L_2-type test based on Pearson's r (Schott (2005));
CJ: the maximum-type test based on Pearson's r (Cai and Jiang (2011)).

5.1. *Computational aspects.* Throughout this section, $\{z_i = (z_{i,1}, z_{i,2})^\top\}_{i\in[n]}$ is a bivariate sample that contains no tie. We first discuss how to compute the U-statistics $\widehat{D}$, $\widehat{R}$ and $\widehat{\tau}^*$ for Hoeffding's D, Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^*, respectively. As we review below, efficient algorithms are available for $\widehat{D}$ and $\widehat{\tau}^*$. The value of $\widehat{R}$ may then be found using the relation in (3.2).

Hoeffding (1948) himself observed that $\widehat{D}$ can be computed in $O(n \log n)$ time via the following formula:

$$\frac{\widehat{D}}{30} = \frac{P - 2(n-2)Q + (n-2)(n-3)S}{n(n-1)(n-2)(n-3)(n-4)}.$$

Here,

$$P := \sum_{i=1}^{n} (r_i - 1)(r_i - 2)(s_i - 1)(s_i - 2),$$

$$Q := \sum_{i=1}^{n} (r_i - 2)(s_i - 1)c_i, \qquad S := \sum_{i=1}^{n} c_i(c_i - 1),$$

and r_i and s_i are the ranks of $z_{i,1}$ among $\{z_{1,1}, \dots, z_{n,1}\}$ and $z_{i,2}$ among $\{z_{1,2}, \dots, z_{n,2}\}$, respectively. Moreover, c_i is the number of pairs $\boldsymbol{z}_{i'}$ for which $z_{i',1} < z_{i,1}$ and $z_{i',2} < z_{i,2}$.

Weihs, Drton and Leung (2016) and Heller and Heller (2016) proposed algorithms for efficient computation of the Bergsma–Dassios–Yanagimoto statistic $\widehat{\tau}^*$. Without loss of generality, let $z_{1,1} < \cdots < z_{n,1}$, that is, $r_i = i$. Weihs, Drton and Leung (2016) proved that $2\widehat{\tau}^*/3 = N_c/\binom{n}{4} - 1/3$ with

$$N_c = \sum_{3 \le \ell < \ell' \le n} \binom{\mathbf{B}_<[\ell, \ell']}{2} + \binom{\mathbf{B}_>[\ell, \ell']}{2},$$

where for all $\ell < \ell'$,

$$\mathbf{B}_<[\ell, \ell'] := \#\{i : i \in [\ell - 1], z_{i,2} < \min(z_{\ell,2}, z_{\ell',2})\} \quad \text{and}$$
$$\mathbf{B}_>[\ell, \ell'] := \#\{i : i \in [\ell - 1], z_{i,2} > \max(z_{\ell,2}, z_{\ell',2})\}.$$

Weihs, Drton and Leung (2016) went on to give an algorithm to compute these counts, and thus $\widehat{\tau}^*$, in $O(n^2 \log n)$ time with little memory use. Heller and Heller (2016) showed that the computation time can be further lowered to $O(n^2)$ via calculation of the following matrix based on the empirical distribution of the ranks r_i and s_i:

$$\mathbf{B}[r, s] := \sum_{i=1}^{n} \mathbb{1}(r_i \le r, s_i \le s), \quad 0 \le r, s \le n.$$

Here, $\mathbf{B}[r, 0] := 0$ and $\mathbf{B}[0, s] := 0$. We may then find $\mathbf{B}_<[\ell, \ell'] = \mathbf{B}[\ell - 1, \min(s_\ell, s_{\ell'}) - 1]$ and $\mathbf{B}_>[\ell, \ell'] = \ell - \mathbf{B}[\ell, \max(s_\ell, s_{\ell'})]$ for all $\ell < \ell'$; recall that s_i is the rank of $z_{i,2}$ in $\{z_{1,2}, \dots, z_{n,2}\}$. As a consequence, formula (3.2) now also yields an $O(n^2)$ algorithm for $\widehat{R}$.

Regarding other competing statistics, note that Pearson's r and Spearman's ρ can be naively computed in time $O(n)$ and $O(n \log n)$, respectively. Knight (1966) proposed an efficient algorithm for computing Kendall's τ that has time complexity $O(n \log n)$. Finally, the algorithm of Huo and Székely (2016) computes the distance covariance statistic in $O(n \log n)$ time.

Table 1 shows empirical computation times for the considered statistics on 1000 bivariate samples of size $n = 100, 200, 400$, and 800, respectively, randomly generated as i.i.d. standard bivariate normal. The timings are based on available functions in `R`. Pearson's r and Spearman's ρ were computed using the basic `cor()` function, with option `method="spearman"` for ρ. Kendall's τ was computed with the function `cor.fk()` from package `pcaPP`, Hoeffding's D with `hoeffD()` from `SymRC`, Bergsma–Dassios–Yanagimoto's τ^* with `tStar()` from `TauStar`, and the distance covariance with `dcov2d()` from `energy`. Blum–Kiefer–Rosenblatt's $\widehat{R}$ was then obtained using identity

TABLE 1
A comparison of computation time for all the correlation statistics considered. The computation time here is the averaged elapsed time (*in milliseconds*) *of* 1000 *replicates of a single experiment*

n	Hoeffding's D	BDY's τ^*	Pearson's r	Spearman's ρ	Kendall's τ	Distance correlation
100	0.270	0.167	0.060	0.121	0.064	0.667
200	0.962	0.543	0.080	0.144	0.085	1.194
400	4.419	2.364	0.099	0.206	0.106	2.313
800	9.683	20.860	0.103	0.327	0.148	4.410

(3.2), and its computation time is thus omitted. All experiments are conducted on a laptop with a 2.6 GHz Intel Core i5 processor and a 8 GB memory.

While the above statistics can all be computed efficiently using special purpose algorithms, our theory also covers general rank-based statistics for which only a naive algorithm that follows the U-statistic definition may be available. The complexity of computing the statistic could then be a high degree polynomial of the sample size. We note that in this case, it may become necessary to use resampling and subsampling techniques to decrease computational effort, as was done by Bergsma and Dassios (2014), Section 4, when applying their statistics before efficient algorithms for its computation were developed.

5.2. *Simulation results*. We evaluate the empirical sizes and powers of the eleven competing tests introduced above for both Gaussian and non-Gaussian distributions. The values reported below are based on 5000 simulations at the nominal significance level of 0.05, with sample size $n \in \{100, 200\}$ and dimension $p \in \{50, 100, 200, 400, 800\}$. All data sets are generated as an i.i.d. sample from the distribution specified for the p-dimensional random vector $\boldsymbol{X}$.

We investigate the sizes of the tests in four settings, where $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ has mutually independent entries. In the following, with slight abuse of notation, we write $f(\boldsymbol{v}) = (f(v_1), \ldots, f(v_p))^\top$ for any univariate function $f : \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$.

EXAMPLE 5.1.

(a) $\boldsymbol{X} \sim N_p(0, \mathbf{I}_p)$ (standard Gaussian).
(b) $\boldsymbol{X} = \boldsymbol{W}^{1/3}$ with $\boldsymbol{W} \sim N_p(0, \mathbf{I}_p)$ (light-tailed Gaussian copula).
(c) $\boldsymbol{X} = \boldsymbol{W}^3$ with $\boldsymbol{W} \sim N_p(0, \mathbf{I}_p)$ (heavy-tailed Gaussian copula).
(d) $X_1, \ldots, X_p$ are i.i.d. with a t-distribution with 3 degrees of freedom.

The simulated sizes of the eight rank-based tests are reported in Table 2. Those of the three distribution-dependent tests are given in Table 3. As expected, the tests derived from Gaussianity (SC, CJ) fail to control the size for heavy-tailed distributions. In contrast, the other tests control the size effectively in most circumstances. A slight size inflation is observed for DHS_D at small sample size, which can be addressed using Monte Carlo approximation to set the critical value. In addition, when considering different pairs of (n, p) in Table 2, as long as n and p grow simultaneously, a trend to the nominal level 0.05 is clear; for example, as (n, p) grows from (100, 200) to (200, 400), the empirical size of DHS_D changes from 0.076 to 0.064, that of DHS_R changes from 0.028 to 0.040, and that of DHS_{τ^*} changes from 0.036 to 0.045. These phenomena back up Corollary 4.1, and this trend persists in more simulations as n and p become even larger.

TABLE 2
Empirical sizes of the eight rank-based tests in Example 5.1

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ
100	50	0.070	0.042	0.047	0.054	0.048	0.056	0.037	0.028
	100	0.073	0.035	0.042	0.055	0.047	0.066	0.034	0.021
	200	0.076	0.028	0.036	0.058	0.050	0.059	0.028	0.015
	400	0.084	0.025	0.035	0.054	0.045	0.065	0.025	0.012
	800	0.088	0.021	0.032	0.055	0.049	0.062	0.023	0.008
200	50	0.054	0.042	0.044	0.048	0.044	0.051	0.037	0.034
	100	0.057	0.042	0.044	0.052	0.047	0.052	0.038	0.032
	200	0.059	0.038	0.042	0.052	0.050	0.055	0.037	0.032
	400	0.064	0.040	0.045	0.051	0.048	0.053	0.038	0.027
	800	0.065	0.034	0.040	0.051	0.047	0.055	0.034	0.024

In order to study the power properties of the different statistics, we consider three sets of examples. We remark that, regarding the power, for L_2-type and maximum-type tests, one cannot dominate the other; compare the power analyses in Section 3.3 in Cai, Liu and Xia (2013) and Section 5.2 in Leung and Drton (2018). To reflect this, we consider two sets of examples that focus on relatively sparse settings (modified based on Yao, Zhang and Shao (2018) and Han, Chen and Liu (2017)) but also include a very dense third setup (modified based on Leung and Drton (2018) with an adjustment to dimension as suggested in Cai and Ma (2013), Theorems 1 and 4).

EXAMPLE 5.2.

(a) The data are generated as $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where

$$\boldsymbol{X}_1 = \big(\boldsymbol{\omega}^\top, \sin(2\pi\boldsymbol{\omega})^\top, \cos(2\pi\boldsymbol{\omega})^\top, \sin(4\pi\boldsymbol{\omega})^\top, \cos(4\pi\boldsymbol{\omega})^\top\big)^\top \in \mathbb{R}^{10}$$

with $\boldsymbol{\omega} \sim N_2(0, \mathbf{I}_2)$, and $\boldsymbol{X}_2 \sim N_{p-10}(0, \mathbf{I}_{p-10})$ independent of $\boldsymbol{X}_1$.

(b) The data are generated as $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where

$$\boldsymbol{X}_1 = \big(\boldsymbol{\omega}^\top, \log(\boldsymbol{\omega}^2)^\top\big)^\top \in \mathbb{R}^{10}$$

with $\boldsymbol{\omega} \sim N_5(0, \mathbf{I}_5)$, and $\boldsymbol{X}_2 \sim N_{p-10}(0, \mathbf{I}_{p-10})$ independent of $\boldsymbol{X}_1$.

TABLE 3
Empirical sizes of the three distribution-dependent tests in Example 5.1

n	p	YZS	SC	CJ	YZS	SC	CJ	YZS	SC	CJ	YZS	SC	CJ
		Results for Case (a)			Results for Case (b)			Results for Case (c)			Results for Case (d)		
100	50	0.048	0.051	0.029	0.052	0.052	0.036	0.055	0.210	0.974	0.055	0.081	0.479
	100	0.054	0.052	0.018	0.048	0.047	0.032	0.052	0.206	1.000	0.053	0.083	0.781
	200	0.059	0.051	0.013	0.055	0.055	0.024	0.052	0.207	1.000	0.058	0.089	0.974
	400	0.049	0.049	0.011	0.053	0.051	0.022	0.052	0.210	1.000	0.055	0.089	1.000
	800	0.050	0.045	0.005	0.050	0.048	0.018	0.055	0.222	1.000	0.051	0.092	1.000
200	50	0.050	0.044	0.032	0.050	0.052	0.040	0.054	0.194	0.955	0.050	0.086	0.527
	100	0.049	0.049	0.029	0.049	0.051	0.036	0.048	0.190	1.000	0.052	0.089	0.850
	200	0.053	0.049	0.030	0.052	0.053	0.035	0.055	0.193	1.000	0.050	0.085	0.996
	400	0.051	0.049	0.022	0.050	0.048	0.035	0.050	0.193	1.000	0.050	0.091	1.000
	800	0.050	0.053	0.018	0.051	0.053	0.033	0.052	0.188	1.000	0.049	0.088	1.000

EXAMPLE 5.3.

(a) The data are drawn as $\boldsymbol{X} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ generated as follows: Consider a random matrix $\boldsymbol{\Delta}$ with all but eight random nonzero entries. We select the locations of four nonzero entries randomly from the upper triangle of $\boldsymbol{\Delta}$, each with a magnitude randomly drawn from the uniform distribution in $[0, 1]$. The other four nonzero entries in the lower triangle are determined to make $\boldsymbol{\Delta}$ symmetric. Finally,

$$\mathbf{R}^* = (1+\delta)\mathbf{I}_p + \boldsymbol{\Delta},$$

where $\delta = \{-\lambda_{\min}(\mathbf{I}_p + \boldsymbol{\Delta}) + 0.05\} \cdot \mathbb{1}\{\lambda_{\min}(\mathbf{I}_p + \boldsymbol{\Delta}) \le 0\}$ and $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of the input.

(b) The data are drawn as $\boldsymbol{X} = \sin(2\pi \mathbf{Z}^{1/3}/3)$, where $\mathbf{Z} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ as in (a).

(c) The data are drawn as $\boldsymbol{X} = \sin(\pi \mathbf{Z}^3/4)$, where $\mathbf{Z} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ as in (a).

EXAMPLE 5.4. The data are drawn as $\boldsymbol{X} \sim N_p(0, \mathbf{R}^*)$, where $\mathbf{R}^* = (1-\varrho)\mathbf{I}_p + \varrho \mathbf{J}_p$ with ϱ such that

(a) $\binom{p}{2}(2 \arcsin \varrho/\pi)^2 = p/n$;
(b) $\binom{p}{2}(2 \arcsin \varrho/\pi)^2 = (3/2) \cdot p/n$;
(c) $\binom{p}{2}(2 \arcsin \varrho/\pi)^2 = 2p/n$.

The powers for Examples 5.2–5.4 are reported in Tables 4–6. Several observations stand out. First, throughout the sparse examples, we found that the proposed tests have the highest powers on average. Among the three proposed tests, the power of DHS_D is highest on average, followed by DHS_{τ^*}. Recall, however, that DHS_D can be subject to slight size inflation. Second, focusing on the results in Example 5.2, we note that, as more independent components are added, the power of YZS significantly decreases. This is as expected and indicates that YZS is less powerful in detection of sparse dependences. In addition, both HCL_τ and HCL_ρ perform unsatisfactorily in Example 5.2, indicating that they are powerless in detecting the considered nonlinear, nonmonotone dependences, an observation that was also made in Yao, Zhang and Shao (2018). Fourth, Tables 4 and 5 jointly confirm the intuition that, for sparse alternatives, the proposed maximum-type tests dominate L_2-type ones including both YZS and LD_{τ^*}, especially when p is large. In addition, we note that, under the setting of Example 5.3, the performances of HCL_τ and HCL_ρ are the second best to the proposed consistent rank-based tests, indicating that there exist cases in which simple rank correlation measures like Kendall's τ and Spearman's ρ can still detect aspects of nonlinear nonmonotone dependences. Fifth, under a Gaussian parametric model, Table 5 (the first part) shows that CJ, the maximum-type test based on Pearson's r, indeed outperforms all others, though the difference between it and the proposed rank-based ones is small. Lastly, Table 6 shows that, as the signals are rather dense, L_2-type tests dominate the maximum-type ones, confirming the intuition and also the theoretical findings that L_2-type ones are more powerful in the dense setting.

We end this section with a discussion of the simulation-based approach. In view of Proposition 2.1, the distributions of rank-based test statistics are invariant to the generating distribution, and hence we may use simulations to approximate the exact distribution of

$$S := \frac{n-1}{\lambda_1 \binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4 \log p - (\mu_1 - 2) \log \log p + \frac{\Lambda}{\lambda_1}.$$

In detail, we pick a large integer M to be the number of independent replications. For each $t \in [M]$, compute $S^{(t)}$ as the value of S for an $n \times p$ data matrix $\mathbf{X}^{(t)} \in \mathbb{R}^{n \times p}$ drawn as having i.i.d. Uniform(0,1) entries. Let $\widehat{F}_{n,p;M}(y) = \frac{1}{M} \sum_{t=1}^{M} \mathbb{1}\{S^{(t)} \le y\}$, $y \in \mathbb{R}$, be the resulting

TABLE 4
Empirical powers of the eleven competing tests in Example 5.2

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 5.2(a)						
100	50	1.000	1.000	1.000	0.058	0.049	1.000	0.089	0.033	0.442	0.047	0.024
	100	1.000	1.000	1.000	0.055	0.045	1.000	0.070	0.025	0.156	0.049	0.018
	200	1.000	1.000	1.000	0.052	0.046	1.000	0.049	0.017	0.071	0.048	0.011
	400	1.000	1.000	1.000	0.058	0.049	0.973	0.043	0.014	0.057	0.050	0.011
	800	1.000	0.827	1.000	0.061	0.052	0.520	0.029	0.009	0.054	0.050	0.007
200	50	1.000	1.000	1.000	0.053	0.045	1.000	0.099	0.038	0.955	0.053	0.033
	100	1.000	1.000	1.000	0.055	0.051	1.000	0.080	0.038	0.435	0.050	0.032
	200	1.000	1.000	1.000	0.048	0.045	1.000	0.060	0.028	0.142	0.045	0.023
	400	1.000	1.000	1.000	0.052	0.047	1.000	0.049	0.023	0.078	0.048	0.023
	800	1.000	1.000	1.000	0.057	0.052	1.000	0.044	0.020	0.053	0.050	0.021
						Results for Example 5.2(b)						
100	50	1.000	1.000	1.000	0.065	0.049	1.000	0.106	0.037	0.984	0.049	0.026
	100	1.000	1.000	1.000	0.054	0.046	1.000	0.078	0.026	0.660	0.046	0.020
	200	1.000	1.000	1.000	0.059	0.052	1.000	0.055	0.018	0.266	0.051	0.014
	400	1.000	1.000	1.000	0.059	0.052	0.996	0.039	0.014	0.107	0.046	0.010
	800	1.000	0.897	1.000	0.059	0.051	0.642	0.030	0.007	0.067	0.052	0.005
200	50	1.000	1.000	1.000	0.062	0.053	1.000	0.120	0.042	1.000	0.050	0.033
	100	1.000	1.000	1.000	0.053	0.047	1.000	0.087	0.040	0.996	0.045	0.036
	200	1.000	1.000	1.000	0.051	0.047	1.000	0.061	0.030	0.729	0.045	0.023
	400	1.000	1.000	1.000	0.053	0.050	1.000	0.050	0.023	0.272	0.053	0.023
	800	1.000	1.000	1.000	0.047	0.044	1.000	0.042	0.021	0.102	0.046	0.016

empirical distribution function. For a specified significance level $\alpha \in (0, 1)$, we may now use the simulated quantile $\widehat{Q}_{\alpha,n,p;M} := \inf\{y \in \mathbb{R} : \widehat{F}_{n,p;M}(y) \geq 1 - \alpha\}$ to form the test

$$\mathsf{T}_\alpha^{\text{exact}} := \mathbb{1}\left\{\frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} > \widehat{Q}_{\alpha,n,p;M}\right\}.$$

The test becomes exact in the large M limit, immediately by the Dvoretzky–Kiefer–Wolfowitz inequality for empirical distribution functions (e.g., Kosorok (2008), Theorem 11.6), and is shown explicitly in the following proposition.

PROPOSITION 5.1. *Under the independence hypothesis H_0, for each (n, p), we have with probability at least $1 - 2/M^2$ that*

$$\sup_{\alpha\in[0,1]} \left|\mathbb{P}\left[S > \widehat{Q}_{\alpha,n,p;M} \mid \{\mathbf{X}^{(t)}\}_{t=1}^M\right] - \left\{1 - \widehat{F}_{n,p;M}(\widehat{Q}_{\alpha,n,p;M})\right\}\right| \leq \left(\frac{\log M}{M}\right)^{1/2}.$$

Table A.1 in the Supplementary Material gives the sizes and powers of the proposed tests with simulation-based critical values ($M = 5000$). The table shows results only for Examples 5.1, 5.3 and 5.4 as the simulated powers under Example 5.2 were all perfectly one. It can be observed that all sizes are now well controlled, with powers of the proposed tests only slightly different from the ones without using simulation. An alternative to the simulation-based approach would be a permutation-based approach, but we find simulation based on the pivotal null distribution simpler to analyze and with the advantage that approximation errors can be made arbitrarily small via larger Monte Carlo samples.

TABLE 5
Empirical powers of the eleven competing tests in Example 5.3

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 5.3(a)						
100	50	0.967	0.962	0.964	0.705	0.586	0.946	0.970	0.966	0.555	0.624	0.973
	100	0.959	0.952	0.954	0.392	0.259	0.914	0.960	0.956	0.252	0.283	0.962
	200	0.950	0.938	0.942	0.161	0.107	0.840	0.950	0.943	0.109	0.115	0.950
	400	0.936	0.924	0.928	0.089	0.064	0.727	0.938	0.931	0.064	0.073	0.941
	800	0.931	0.911	0.918	0.061	0.049	0.539	0.929	0.916	0.051	0.051	0.931
200	50	0.991	0.991	0.991	0.912	0.891	0.988	0.993	0.992	0.871	0.906	0.993
	100	0.984	0.985	0.985	0.728	0.627	0.974	0.988	0.987	0.579	0.650	0.989
	200	0.984	0.983	0.983	0.408	0.278	0.954	0.987	0.985	0.255	0.299	0.988
	400	0.986	0.983	0.983	0.166	0.110	0.917	0.986	0.985	0.111	0.115	0.989
	800	0.980	0.976	0.978	0.073	0.060	0.839	0.983	0.980	0.058	0.063	0.986
						Results for Example 5.3(b)						
100	50	0.759	0.642	0.687	0.244	0.167	0.623	0.623	0.553	0.277	0.260	0.786
	100	0.747	0.624	0.670	0.131	0.091	0.555	0.607	0.540	0.131	0.125	0.758
	200	0.720	0.583	0.635	0.082	0.062	0.444	0.578	0.502	0.080	0.075	0.714
	400	0.702	0.557	0.615	0.065	0.054	0.333	0.549	0.471	0.060	0.061	0.678
	800	0.679	0.512	0.577	0.057	0.048	0.218	0.517	0.431	0.052	0.051	0.638
200	50	0.897	0.843	0.866	0.423	0.343	0.825	0.810	0.767	0.577	0.550	0.928
	100	0.880	0.819	0.846	0.248	0.170	0.753	0.784	0.732	0.287	0.273	0.912
	200	0.855	0.789	0.818	0.128	0.088	0.670	0.757	0.714	0.129	0.128	0.891
	400	0.849	0.768	0.799	0.074	0.059	0.571	0.743	0.689	0.065	0.064	0.875
	800	0.820	0.738	0.772	0.051	0.045	0.450	0.713	0.654	0.053	0.051	0.852
						Results for Example 5.3(c)						
100	50	0.654	0.579	0.608	0.209	0.137	0.541	0.582	0.513	0.111	0.106	0.365
	100	0.656	0.566	0.599	0.109	0.072	0.464	0.580	0.502	0.071	0.064	0.344
	200	0.635	0.527	0.571	0.069	0.055	0.364	0.539	0.455	0.056	0.051	0.311
	400	0.617	0.496	0.546	0.068	0.059	0.256	0.516	0.421	0.053	0.058	0.277
	800	0.597	0.455	0.507	0.055	0.049	0.164	0.487	0.370	0.055	0.049	0.238
200	50	0.824	0.789	0.803	0.396	0.302	0.750	0.785	0.753	0.238	0.211	0.606
	100	0.812	0.773	0.788	0.219	0.143	0.681	0.768	0.732	0.113	0.100	0.570
	200	0.792	0.752	0.767	0.101	0.072	0.596	0.750	0.711	0.063	0.059	0.543
	400	0.776	0.728	0.744	0.070	0.054	0.499	0.730	0.689	0.058	0.057	0.513
	800	0.755	0.699	0.723	0.052	0.048	0.360	0.699	0.646	0.044	0.051	0.473

6. Discussion.

6.1. *Discussion of Assumption* 2.1. Assumption 2.1 plays a key role in our analysis. It synthesizes crucial properties satisfied by the three rank correlation statistics from Examples 2.1–2.3.

From a more general perspective, one might ask whether there is an exact relation between Assumption 2.1 and the properties of I- and D-consistency summarized in Weihs, Drton and Meinshausen (2018). As a matter of fact, to our knowledge, most existing test statistics (including rank-based, distance covariance-based, and kernel-based ones) that permit consistent assessment of pairwise independence are asymptotically equivalent to U-statistics with the corresponding kernels degenerate under the null, which echoes Assumption 2.1(ii). The only exception is a new rank correlation measure that was just proposed (Chatterjee (2020)), whose limiting distribution is normal. Its analysis uses the permutation theory and, in particular, is not based on the U-statistic framework. Assumption 2.1(iii), on the other hand, is much more

TABLE 6
Empirical powers of the eleven competing tests in Example 5.4

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 5.4(a)						
100	50	0.102	0.068	0.074	0.532	0.524	0.350	0.062	0.046	0.474	0.578	0.042
	100	0.104	0.056	0.066	0.578	0.560	0.361	0.052	0.036	0.492	0.620	0.033
	200	0.096	0.035	0.048	0.583	0.565	0.343	0.037	0.022	0.488	0.620	0.018
	400	0.104	0.040	0.050	0.542	0.534	0.320	0.038	0.018	0.471	0.610	0.012
	800	0.095	0.018	0.032	0.570	0.552	0.344	0.027	0.007	0.487	0.620	0.005
200	50	0.104	0.080	0.086	0.564	0.544	0.357	0.081	0.072	0.478	0.614	0.068
	100	0.073	0.052	0.059	0.590	0.580	0.357	0.054	0.043	0.509	0.654	0.052
	200	0.085	0.061	0.064	0.594	0.585	0.336	0.052	0.040	0.488	0.652	0.040
	400	0.075	0.040	0.049	0.604	0.591	0.332	0.038	0.028	0.498	0.668	0.024
	800	0.067	0.036	0.044	0.586	0.573	0.320	0.034	0.027	0.488	0.640	0.026
						Results for Example 5.4(b)						
100	50	0.130	0.078	0.086	0.792	0.782	0.554	0.076	0.064	0.722	0.836	0.055
	100	0.110	0.056	0.062	0.808	0.800	0.584	0.052	0.035	0.746	0.848	0.032
	200	0.099	0.046	0.060	0.810	0.800	0.553	0.042	0.026	0.738	0.850	0.021
	400	0.110	0.030	0.041	0.808	0.797	0.587	0.034	0.014	0.738	0.854	0.012
	800	0.098	0.020	0.033	0.816	0.804	0.579	0.023	0.008	0.745	0.872	0.006
200	50	0.116	0.094	0.098	0.802	0.801	0.546	0.103	0.084	0.718	0.858	0.098
	100	0.098	0.072	0.076	0.827	0.822	0.571	0.075	0.062	0.768	0.878	0.058
	200	0.063	0.040	0.042	0.848	0.840	0.570	0.036	0.030	0.764	0.888	0.030
	400	0.070	0.048	0.055	0.834	0.829	0.578	0.042	0.032	0.752	0.883	0.030
	800	0.081	0.036	0.046	0.866	0.862	0.560	0.041	0.028	0.788	0.907	0.030
						Results for Example 5.4(c)						
100	50	0.157	0.102	0.116	0.904	0.900	0.731	0.093	0.069	0.864	0.926	0.076
	100	0.124	0.067	0.082	0.914	0.909	0.738	0.058	0.036	0.878	0.943	0.042
	200	0.115	0.051	0.059	0.918	0.913	0.748	0.046	0.028	0.880	0.947	0.018
	400	0.112	0.034	0.046	0.930	0.926	0.738	0.038	0.017	0.888	0.954	0.009
	800	0.101	0.030	0.039	0.927	0.924	0.744	0.029	0.012	0.879	0.946	0.012
200	50	0.120	0.100	0.098	0.935	0.932	0.740	0.110	0.098	0.894	0.952	0.118
	100	0.107	0.082	0.085	0.941	0.939	0.740	0.072	0.066	0.892	0.960	0.065
	200	0.096	0.062	0.072	0.962	0.960	0.768	0.064	0.048	0.930	0.976	0.046
	400	0.077	0.042	0.046	0.964	0.962	0.792	0.037	0.028	0.930	0.978	0.024
	800	0.090	0.043	0.054	0.956	0.956	0.776	0.044	0.028	0.922	0.980	0.016

specific and related to the particular properties of rank-based consistent tests. This assumption, however, is key to the establishment of Theorem 4.2.

6.2. *Discussion of* τ^*. In this section, we give new perspectives on Bergsma–Dassios–Yanagimoto's correlation measure $\tau^* := \mathbb{E}h_{\tau^*}$, introduced in Example 2.3. Hoeffding (1948) stated a problem about the relationship between equiprobable rankings and independence that was solved by Yanagimoto (1970). In the proof of his Proposition 9, Yanagimoto (1970) presented a correlation measure that is proportional to τ^* of Bergsma–Dassios if the pair is absolutely continuous. Accordingly, we term the correlation "Bergsma–Dassios–Yanagimoto's τ^*". Yanagimoto's key relation gives rise to an interesting identity between Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^* statistics. This identity appears to be unknown in the literature. In detail, if $z_1, \dots, z_6 \in \mathbb{R}^2$ have no tie among their first and their second entries, respectively, then

$$
\begin{aligned}
(6.1)\quad & 3 \cdot \binom{6}{5}^{-1} \sum_{1 \le i_1 < \cdots < i_5 \le 6} h_D(z_{i_1}, \ldots, z_{i_5}) + 2h_R(z_1, \ldots, z_6) \\
& = 5 \cdot \binom{6}{4}^{-1} \sum_{1 \le i_1 < \cdots < i_4 \le 6} h_{\tau^*}(z_{i_1}, \ldots, z_{i_4}).
\end{aligned}
$$

Equation (6.1) can be easily verified by calculating all 6! entrywise permutations of $\{1, 2, \ldots, 6\}$, but may be false when ties exist. Using the identity, we can make a step towards proving the conjecture raised in Bergsma and Dassios (2014), that is, for an arbitrary random pair $(Z_1, Z_2)^\top \in \mathbb{R}^2$, do we have $\mathbb{E}h_{\tau^*} \ge 0$ with equality if and only if Z_1 and Z_2 are independent?

THEOREM 6.1. *For any random vector $\mathbf{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$ with continuous marginal distributions, we have $\mathbb{E}h_{\tau^*} \ge 0$ and the equality holds if and only if Z_1 is independent of Z_2.*

Similarly, a monotonicity property of $\mathbb{E}h_D$ and $\mathbb{E}h_R$ proved by Yanagimoto (1970), Section 2, extends to $\mathbb{E}h_{\tau^*}$. We state the Gaussian version of this property.

THEOREM 6.2. *If $\mathbf{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$ is bivariate Gaussian with (Pearson) correlation ρ, then $\mathbb{E}h_D$ and $\mathbb{E}h_R$, and, thus, also $\mathbb{E}h_{\tau^*}$ are increasing functions of $|\rho|$.*

Theorem 6.1 complements the results in Theorem 1 in Bergsma and Dassios (2014) to include random vectors with continuous margins and a bivariate joint distribution that is continuous (implied by marginal continuity) but need not be absolutely continuous. Such an example of distribution on $\mathbb{R}^2$ that has continuous margins but is not absolutely continuous has been constructed in Remark 1 in Yanagimoto (1970), where it is used to illustrate an inconsistency problem about Hoeffding's D. A simpler example is the uniform distribution on the unit circle in $\mathbb{R}^2$. For this, we revisit a comment of Weihs, Drton and Meinshausen (2018) who noted that based on existing literature "it is not guaranteed that $\mathbb{E}h_{\tau^*} > 0$ when $(X, Y)^\top$ is generated uniformly on the unit circle in $\mathbb{R}^2$." We are able to calculate the values of D and R for this example, and thus, can deduce the value of τ^*.

PROPOSITION 6.1. *For $(X, Y)^\top$ following the uniform distribution on the unit circle in $\mathbb{R}^2$, we have $\mathbb{E}h_D = \mathbb{E}h_R = \mathbb{E}h_{\tau^*} = 1/16$.*

Acknowledgments. The authors thank three anonymous referees, the anonymous Associate Editor, and the Editor, Richard J. Samworth, for their constructive comments and suggestions, which helped to improve the quality of this paper.

This work was partially supported by NSF Grants DMS-1712535 and DMS-1712536.

SUPPLEMENTARY MATERIAL

Supplement to "High-dimensional consistent independence testing with maxima of rank correlations" (DOI: 10.1214/19-AOS1926SUPP; .pdf). This supplement contains all the technical proofs and additional simulation results.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR1990662

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542. MR1235426

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694

BAO, Z. (2019). Tracy–Widom limit for Kendall's tau. *Ann. Statist.* **47** 3504–3532. MR4025750 https://doi.org/10.1214/18-AOS1786

BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 1–32. MR2988403 https://doi.org/10.1214/EJP.v17-1962

BENTKUS, V. and GÖTZE, F. (1997). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Related Fields* **109** 367–416. MR1481126 https://doi.org/10.1007/s004400050136

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

CHATTERJEE, S. K. (2020). A new coefficient of correlation. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2020.1758115

DRTON, M., HAN, F. and SHI, H. (2020). Supplement to "High-dimensional consistent independence testing with maxima of rank correlations." https://doi.org/10.1214/19-AOS1926SUPP

GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

GÖTZE, F. and ZAITSEV, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *Ann. Probab.* **42** 354–397. MR3161488 https://doi.org/10.1214/13-AOP839

GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 20 (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) 984–991. Curran Associates, Red Hook, NY.

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

HAN, F., XU, S. and ZHOU, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli* **24** 1787–1833. MR3757515 https://doi.org/10.3150/16-BEJ912

HELLER, R., HELLER, Y., KAUFMAN, S., BRILL, B. and GORFINE, M. (2016). Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17** Paper No. 29, 54. MR3491123

HELLER, Y. and HELLER, R. (2016). Computing the Bergsma Dassios sign-covariance. Available at arXiv:1605.08732.

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

HOLLANDER, M., WOLFE, D. A. and CHICKEN, E. (2014). *Nonparametric Statistical Methods*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3221959

HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. MR3556612 https://doi.org/10.1080/00401706.2015.1054435

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857 https://doi.org/10.1214/13-AOS1134

KENDALL, M. and STUART, A. (1979). *The Advanced Theory of Statistics. Vol.* 2: *Inference and Relationship*, 4th ed. Charles Griffin & Co. Ltd., London.

KINNEY, J. B. and ATWAL, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **111** 3354–3359. MR3200177 https://doi.org/10.1073/pnas.1309933111

KNIGHT, W. R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *J. Amer. Statist. Assoc.* **61** 436–439.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

NANDY, P., WEIHS, L. and DRTON, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Stat.* **10** 2287–2311. MR3541972 https://doi.org/10.1214/16-EJS1166

PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York. MR0092296

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197 https://doi.org/10.1093/biomet/92.4.951

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

WEIHS, L., DRTON, M. and LEUNG, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Comput. Statist.* **31** 315–328. MR3481807 https://doi.org/10.1007/s00180-015-0639-x

WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

YANAGIMOTO, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.* **22** 57–63.

YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. MR2327033 https://doi.org/10.1090/S0002-9947-07-04192-X