Distribution-Free Consistent Tests of Independence via Marginal and Multivariate Ranks

Hongjian Shi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Fang Han, Chair

Mathias Drton

Yanqin Fan

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Distribution-Free Consistent Tests of Independence
via Marginal and Multivariate Ranks

Hongjian Shi

Chair of the Supervisory Committee:
Professor Fang Han
Statistics

Testing independence is a fundamental statistical problem that has received much attention in literature. In this dissertation, we consider testing independence under two different settings. The first is testing mutual independence of many covariates, and the second is testing independence of two random vectors. For both settings, we propose, for the first time, distribution-free and consistent tests of independence via marginal or multivariate ranks. Moreover, we establish the optimal efficiency in the statistical sense of both tests. In addition, we also investigate the power of a simple consistent rank correlation coefficient recently proposed by Chatterjee (2021) against local alternatives. Our results show that Chatterjee's coefficient is unfortunately statistically inefficient.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number | Page

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Prof. Fang Han, and I would like to thank Prof. Mathias Drton, who is like my second advisor. I am grateful for their mentorship throughout my doctoral study at the University of Washington. The thesis would be very different without their invaluable guidance and input.

I would also like to thank my committee members, Prof. Yanqin Fan and Prof. Jon Wellner, for serving on my committee and offering their very thoughtful advice, which helped me improve my thesis. I would like to extend my gratitude to my wonderful collaborator, Prof. Marc Hallin at Université Libre de Bruxelles.

Last but not least, I would like to thank my family and friends, who always encourage and support me. I am lucky to have all of you in my life.

DEDICATION

to my family

Chapter 1

INTRODUCTION

Rank correlations have found many innovative applications in the last decade. In particular, suitable rank correlations have been used for consistent tests of independence. In this dissertation, we consider testing independence via marginal and multivariate ranks under two different settings. We will study testing mutual independence for high-dimensional observations in Chapter 2, and move on to the problem of testing independence between two random vectors/variables in Chapters 3 and 4.

Chapter 2 is concerned with testing mutual independence among all entries in a random vector based on finite observations. Popular tests based on linear and simple rank correlations are known to be incapable of detecting non-linear, non-monotone 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.

In Chapter 3, we consider testing independence between two random vectors. When it reduces to the univariate case (both random vectors are in dimension one), using ranks is especially appealing for continuous data as tests become distribution-free. However, the traditional concept of ranks relies on ordering data and is, thus, tied to univariate observations. As a result, it has long remained unclear how one may construct distribution-free yet consistent tests of independence between random vectors. In this chapter, we address this problem by laying out a general framework for designing dependence measures that give tests of multivariate independence that are not only consistent and distribution-free but which we also prove to be statistically efficient. Our framework leverages the recently introduced concept of center-outward ranks and signs, a multivariate generalization of traditional ranks, and adopts a common standard form for dependence measures that encompasses many popular examples. In a unified study, we derive a general asymptotic representation of center-outward rank-based test statistics under independence, extending to the multivariate setting the classical Hájek asymptotic representation results. This representation permits direct calculation of limiting null distributions and facilitates a local power analysis that provides strong support for the center-outward approach by establishing, for the first time, the nontrivial power of center-outward rank-based tests over root-n neighborhoods within the class of quadratic mean differentiable alternatives.

In Chapter 4, we focus on the problem of testing independence between two univariate random variables. Chatterjee (2021) introduced a simple new rank correlation coefficient that has attracted much recent attention. 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 chapter compares Chatterjee's new correlation coefficient to 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 contrast their computational efficiency 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 sub-optimal compared to D, R, and τ^*. The situation is more subtle for a related earlier estimator of Dette et al. (2013). These results favor D, R, and τ^* over Chatterjee's new correlation coefficient for the purpose of testing independence.

The main contents of this thesis are taken from the following articles and manuscripts with minor modification. Chapter 2 is adapted from "High-dimensional consistent independence testing with maxima of rank correlations", coauthored with Mathias Drton and Fang Han, published on The Annals of Statistics (Drton et al., 2020). Chapter 3 is drawn from "On universally consistent and fully distribution-free rank tests of vector independence", coauthored with Marc Hallin, Mathias Drton and Fang Han (Shi et al., 2020); it extends an earlier paper "Distribution-free consistent independence tests via center-outward ranks and signs", coauthored with Mathias Drton and Fang Han, accepted to the Journal of the American Statistical Association (Shi et al., 2021a). The last chapter is from "On the power of Chatterjee's rank correlation" coauthored with Mathias Drton and Fang Han, accepted to Biometrika (Shi et al., 2021b).

Chapter 2

HIGH-DIMENSIONAL CONSISTENT INDEPENDENCE TESTING WITH MAXIMA OF RANK CORRELATIONS

2.1 Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ 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

$$H_0 : X_1, \ldots, X_p \text{ are mutually independent,} \tag{2.1.1}$$

based on n independent realizations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of $\boldsymbol{X}$. Testing H_0 is a core problem in multivariate statistics that has attracted the attention of statisticians for decades; see e.g. the exposition in Anderson (2003, Chap. 9) or Muirhead (1982, Chap. 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 et al. (2012), Gao et al. (2017), Han et al. (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 et al. (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.

The aforementioned approaches are largely built on linear and simple rank correlations. These, however, are incapable of detecting more complicated non-linear, non-monotone 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 et al., 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 et al. (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 et al. (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 et al. (2017), Leung and Drton (2018), and Yao et al. (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 et al., 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 et al. (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 2.2, 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 2.3 have a null distribution that converges to a (nonstandard) Gumbel distribution under high-dimensional asymptotics. This is in contrast to the results in Han et al. (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 2.4.

The theoretical advantages of our tests are highlighted in simulation studies (Section 2.5). 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 towards proving Bergsma–Dassios's conjecture about general D-consistency of τ^*. This observation, along with other discussions, is given in Section 2.6. All proofs and additional simulation results are deferred to a supplement.

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 $[\![m]\!] = \{1, 2, \ldots, m\}$ and write $\mathcal{P}_m$ for the set of all $m!$ permutations of $[\![m]\!]$. Let $\mathbf{M} = [M_{jk}] \in \mathbb{R}^{p\times p}$, and I, J be two subsets of $[\![p]\!]$. Then both $\mathbf{M}_{I,J}$ and $\mathbf{M}[I, J]$ are used to refer to the sub-matrix of $\mathbf{M}$ with rows indexed by I and columns indexed by J. The matrix $\mathrm{diag}(\mathbf{M}) \in \mathbb{R}^{p\times p}$ 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 from line to line.

2.2 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 (2.1.1). 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 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent copies of $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$, with $\boldsymbol{X}_i = (X_{1i}, \ldots, X_{pi})^\top$. Let $j \neq k \in [\![p]\!]$, 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:

$$\widehat{U}_{jk} = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} h\left\{ \binom{X_{ji_1}}{X_{ki_1}}, \ldots, \binom{X_{ji_m}}{X_{ki_m}} \right\}. \tag{2.2.1}$$

For our purposes, the kernel h may always be assumed to be *symmetric*, i.e., $h(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m) = h(\boldsymbol{z}_{\sigma(1)}, \ldots, \boldsymbol{z}_{\sigma(m)})$ for all permutations $\sigma \in \mathcal{P}_m$ and $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathbb{R}^2$. Letting $\boldsymbol{z}_i = (z_{1i}, z_{2i})^\top$, if both vectors $(z_{11}, \ldots, z_{1m})$ and $(z_{21}, \ldots, z_{2m})$ are free of ties, i.e., have marginal distinct entries, then we have well-defined vectors of ranks $(r_{11}, \ldots, r_{1m})$ and $(r_{21}, \ldots, r_{2m})$, and we define $\boldsymbol{r}_i = (r_{1i}, r_{2i})^\top$ for $1 \leq i \leq n$. Now a kernel is *rank-based* if

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

for all $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathbb{R}^2$ with $(z_{11}, \ldots, z_{1m})$ and $(z_{21}, \ldots, z_{2m})$ free of ties. In this case, we also say that the "correlation" statistic $\widehat{U}_{jk}$ as well as the corresponding "correlation measure" $\mathrm{E}\widehat{U}_{jk}$ 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, e.g., Chapter 31 in Kendall and Stuart (1979), Lemma C4 in the supplement of Han et al. (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 2.2.1. *Under the null hypothesis in* (2.1.1) *and assuming continuous margins, we have:*

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

(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 $\mathrm{P}_{\boldsymbol{Z}}$, we write

$$h_\ell(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) := \mathrm{E}h(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell, \boldsymbol{Z}_{\ell+1}, \ldots, \boldsymbol{Z}_m)$$

and

$$\begin{aligned} &h^{(\ell)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) \\ &:= h_\ell(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) - \mathrm{E}h - \sum_{k=1}^{\ell-1} \sum_{1 \leq i_1 < \cdots < i_k \leq \ell} h^{(k)}(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_k}; \mathrm{P}_{\boldsymbol{Z}}), \end{aligned} \tag{2.2.2}$$

where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are m independent random vectors with distribution $\mathrm{P}_{\boldsymbol{Z}}$ and $\mathrm{E}h := \mathrm{E}h(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m)$. The kernel as well as the corresponding U-statistic is *degenerate* under $\mathrm{P}_{\boldsymbol{Z}}$ if $h_1(\cdot)$ has variance zero. We use the term *completely degenerate* to indicate that the variances of $h_1(\cdot), \ldots, h_{m-1}(\cdot)$ are all zero. Finally, let P_0 be the uniform distribution on $[0,1]$, and write $\mathrm{P}_0 \otimes \mathrm{P}_0$ for its product measure, the uniform distribution on $[0,1]^2$. Note that by Proposition 2.2.1(i), the study of $\widehat{U}_{jk}$ under independent continuous margins X_j and X_k can be reduced to the case with $(X_j, X_k)^\top \sim \mathrm{P}_0 \otimes \mathrm{P}_0$.

Assumption 2.2.1. *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, i.e.,* $\mathrm{E}\{h_1(\boldsymbol{Z}_1;\mathrm{P}_0\otimes\mathrm{P}_0)\}^2=0$ *as* $\boldsymbol{Z}_1\sim\mathrm{P}_0\otimes\mathrm{P}_0$.

(iii) $h_2(\boldsymbol{z}_1,\boldsymbol{z}_2;\mathrm{P}_0\otimes\mathrm{P}_0)$ *has uniformly bounded eigenfunctions, that is, it admits the expansion*

$$h_2(\boldsymbol{z}_1,\boldsymbol{z}_2;\mathrm{P}_0\otimes\mathrm{P}_0)=\sum_{v=1}^{\infty}\lambda_v\phi_v(\boldsymbol{z}_1)\phi_v(\boldsymbol{z}_2),$$

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

$$\mathrm{E}h_2(\boldsymbol{z}_1,\boldsymbol{Z}_2)\phi(\boldsymbol{Z}_2)=\lambda\phi(\boldsymbol{z}_1)\quad\textit{for all } \boldsymbol{z}_1\in\mathbb{R}^2,$$

with $\boldsymbol{Z}_2\sim\mathrm{P}_0\otimes\mathrm{P}_0$, $\lambda_1\geq\lambda_2\geq\cdots\geq0$, $\Lambda:=\sum_{v=1}^{\infty}\lambda_v\in(0,\infty)$, *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 2.2.1 (Hoeffding's D)**.** Letting $\boldsymbol{z}_i=(z_{1i},z_{2i})^\top$, from the symmetric kernel

$$\begin{aligned}h_D(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_5):=\frac{1}{16}&\sum_{(i_1,\ldots,i_5)\in\mathcal{P}_5}\\ &\Big[\Big\{\mathbb{1}(z_{1i_1}\leq z_{1i_5})-\mathbb{1}(z_{1i_2}\leq z_{1i_5})\Big\}\Big\{\mathbb{1}(z_{1i_3}\leq z_{1i_5})-\mathbb{1}(z_{1i_4}\leq z_{1i_5})\Big\}\Big]\\ &\Big[\Big\{\mathbb{1}(z_{2i_1}\leq z_{2i_5})-\mathbb{1}(z_{2i_2}\leq z_{2i_5})\Big\}\Big\{\mathbb{1}(z_{2i_3}\leq z_{2i_5})-\mathbb{1}(z_{2i_4}\leq z_{2i_5})\Big\}\Big],\end{aligned}$$

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 $\mathrm{E}h_D$. The kernel $h_D(\cdot)$ satisfies the first two properties in Assumption 2.2.1 in view of the results in Hoeffding (1948). To verify the last property, we note that under the measure $\mathrm{P}_0 \otimes \mathrm{P}_0$, $h_{D,2}(\cdot)$ is known to have eigenvalues

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

see, e.g., Proposition 7 in Weihs et al. (2018) or Theorem 4.4 in Nandy et al. (2016). The corresponding eigenfunctions are

$$\phi_{i,j;D}\{(z_{11}, z_{21})^\top\} = 2\cos(\pi i z_{11})\cos(\pi j z_{21}), \quad i, j \in \mathbb{Z}_+.$$

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 $\mathrm{E}h_D \ge 0$ (Hoeffding, 1948, p. 547). Furthermore, it has been proven that, once the pair is absolutely continuous in $\mathbb{R}^2$, the correlation measure $\mathrm{E}h_D = 0$ 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 2.2.2 (Blum–Kiefer–Rosenblatt's R)**.** The symmetric kernel

$$h_R(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6) := \frac{1}{32} \sum_{(i_1,\ldots,i_6)\in\mathcal{P}_6} \Big[\Big\{\mathbb{1}(z_{1i_1} \le z_{1i_5}) - \mathbb{1}(z_{1i_2} \le z_{1i_5})\Big\}\Big\{\mathbb{1}(z_{1i_3} \le z_{1i_5}) - \mathbb{1}(z_{1i_4} \le z_{1i_5})\Big\}\Big] \Big[\Big\{\mathbb{1}(z_{2i_1} \le z_{2i_6}) - \mathbb{1}(z_{2i_2} \le z_{2i_6})\Big\}\Big\{\mathbb{1}(z_{2i_3} \le z_{2i_6}) - \mathbb{1}(z_{2i_4} \le z_{2i_6})\Big\}\Big]$$

yields Blum–Kiefer–Rosenblatt's R statistic (Blum et al., 1961), which is a rank-based U-statistic of order 6. One can verify the three properties in Assumption 2.2.1 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 $\mathrm{E}h_R \ge 0$ with equality if and only if the pair is independent, and no continuity assumption is needed at all; cf. page 490 of Blum et al. (1961).

Example 2.2.3 (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

$$\begin{aligned} & h_{\tau^*}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4) \\ := & \frac{1}{16} \sum_{(i_1, \ldots, i_4) \in \mathcal{P}_4} \Big\{ \mathbb{1}(z_{1i_1}, z_{1i_3} < z_{1i_2}, z_{1i_4}) + \mathbb{1}(z_{1i_2}, z_{1i_4} < z_{1i_1}, z_{1i_3}) \\ & \qquad - \mathbb{1}(z_{1i_1}, z_{1i_4} < z_{1i_2}, z_{1i_3}) - \mathbb{1}(z_{1i_2}, z_{1i_3} < z_{1i_1}, z_{1i_4}) \Big\} \\ & \qquad \Big\{ \mathbb{1}(z_{2i_1}, z_{2i_3} < z_{2i_2}, z_{2i_4}) + \mathbb{1}(z_{2i_2}, z_{2i_4} < z_{2i_1}, z_{2i_3}) \\ & \qquad - \mathbb{1}(z_{2i_1}, z_{2i_4} < z_{2i_2}, z_{2i_3}) - \mathbb{1}(z_{2i_2}, z_{2i_3} < z_{2i_1}, z_{2i_4}) \Big\}. \end{aligned}$$

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 2.2.1 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 $\mathrm{E}h_{\tau^*} \geq 0$ 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 2.6.2 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^*.

2.3 Maximum-type tests of mutual independence

We now turn to tests of the mutual independence hypothesis H_0 in (2.1.1). As in Han et al. (2017), we propose maximum-type tests. However, in contrast to Han et al. (2017), we suggest the use of consistent and rank-based correlations with the practical choices being the ones from Examples 2.2.1–2.2.3. 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 (2.2.1) 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 2.2.1(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 $X \stackrel{\mathsf{d}}{=} Y$ if two random variables X and Y have the same distribution, and we use $\stackrel{\mathsf{d}}{\longrightarrow}$ 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 et al. (2017), then extreme-value theory combined with Proposition 2.2.1 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 2.2.1(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, Chap. 5.5.2).

Proposition 2.3.1. *Let* $\boldsymbol{X}$ *have continuous margins, and let* $j \neq k$*. If* $h(\cdot)$ *satisfies Assumption 2.2.1, then under* H_0*,*

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

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

Note that the weak convergence result for degenerate U-statistics in Proposition 2.3.1

holds under much weaker conditions than Assumption 2.2.1; see the main theorem in Serfling (1980, Chap. 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 $\{\widehat{U}_{jk}; j < k\}$ are in general not mutually independent.

Proposition 2.3.2. *Let $Y_1, \ldots, Y_d$ be $d = p(p-1)/2$ independent copies of $\zeta \overset{\mathsf{d}}{=} \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1)$. 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

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

where μ_1 is the multiplicity of the largest eigenvalue λ_1 in the sequence $\{\lambda_1, \lambda_2, \ldots\}$, $\kappa := \prod_{v=\mu_1+1}^{\infty}(1 - \lambda_v/\lambda_1)^{-1/2}$, 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 2.3.2, we recover the Gumbel distribution derived by Han et al. (2017). Based on Propositions 2.3.1 and 2.3.2, for any pre-specified significance level $\alpha \in (0, 1)$, our proposed test is

$$\mathsf{T}_\alpha := \mathbb{1}\Big\{ \frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} > Q_\alpha \Big\}, \tag{2.3.1}$$

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 2.4.

Example 2.3.1 ("Extreme D")**.** Hoeffding's D statistic introduced in Example 2.2.1 is

$$\widehat{D}_{jk} := \binom{n}{5}^{-1} \sum_{i_1 < \cdots < i_5} h_D\{(X_{ji_1}, X_{ki_1})^\top, \ldots, (X_{ji_5}, X_{ki_5})^\top\}.$$

According to (2.3.1), the corresponding test is

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

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

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

Example 2.3.2 ("Extreme R")**.** Blum–Kiefer–Rosenblatt's R statistic from Example 2.2.2 is

$$\widehat{R}_{jk} := \binom{n}{6}^{-1} \sum_{i_1 < \cdots < i_6} h_R\{(X_{ji_1}, X_{ki_1})^\top, \ldots, (X_{ji_6}, X_{ki_6})^\top\}.$$

According to (2.3.1), the corresponding test is

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

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

Example 2.3.3 ("Extreme τ^*")**.** Bergsma–Dassios–Yanagimoto's τ^* statistic from Example 2.2.3 is

$$\widehat{\tau}^*_{jk} := \binom{n}{4}^{-1} \sum_{i_1 < \cdots < i_4} h_{\tau^*}\{(X_{ji_1}, X_{ki_1})^\top, \ldots, (X_{ji_4}, X_{ki_4})^\top\}.$$

According to (2.3.1), it yields the test

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

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

Note that, by the definitions of the kernels and the identity (2.6.1) that will be introduced in Section 2.6.2, as long as there is no tie in the data, for any $j, k \in [\![p]\!]$,

$$\widehat{D}_{jj} = \widehat{R}_{jj} = \widehat{\tau}^*_{jj} = 1 \quad \text{and} \quad 3\widehat{D}_{jk} + 2\widehat{R}_{jk} = 5\widehat{\tau}^*_{jk}. \tag{2.3.2}$$

Remark 2.3.1. In applying the above tests we have intrinsically assumed that there are no ties among the entries $X_{j1}, \ldots, X_{jn}$ 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 et al. (2014) for more discussions on how to break ties for rank-based tests.

2.4 Theoretical analysis

This section provides theoretical justifications of the tests proposed in Section 2.3. 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.

2.4.1 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 non-degenerate U-statistics (see, for example, Shao and Zhou (2016) for some recent developments) as well as Berry–Esseen-type bounds for degenerate U-statistics (see Bentkus 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 2.4.1 (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* 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)* $h_1(Z_1; \mathrm{P}_Z) = 0$ *almost surely, and (iii)* $h_2(z_1, z_2; \mathrm{P}_Z)$ *admits the eigenfunction expansion,*

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

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

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

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

$$\theta < \sup \left\{ q \in [0, 1/3) : \sum\nolimits_{v > \lfloor n^{(1-3q)/5} \rfloor} \lambda_v = O(n^{-q}) \right\} \tag{2.4.1}$$

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

In Theorem 2.4.1, 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 2.4.1 and combining it with Proposition 2.3.2, 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 2.3.2. Hence, the proposed test T_α in (2.3.1) can effectively control the size.

Theorem 2.4.2 (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 2.2.1. Define the parameter θ as in (2.4.1). 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*

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

Consequently,

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

where P_{H_0} represents the probability under the null hypothesis H_0.

Note that the proof of Theorem 2.4.2 uses the Chen–Stein method, via Theorem 1 of Arratia et al. (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 et al. (2018).

As a simple consequence of Theorem 2.4.2, the following corollary shows that the tests in Examples 2.3.1–2.3.3 have asymptotically correct sizes, with θ being explicitly calculated.

Corollary 2.4.1. *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 pre-specified constant $\delta > 0$. Then*

$$\begin{aligned}
&\mathrm{P}_{H_0}(\mathsf{T}_{D,\alpha} = 1) = \alpha + o(1), \quad \mathrm{P}_{H_0}(\mathsf{T}_{R,\alpha} = 1) = \alpha + o(1), \\
\text{and} \quad &\mathrm{P}_{H_0}(\mathsf{T}_{\tau^*,\alpha} = 1) = \alpha + o(1).
\end{aligned}$$

2.4.2 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 (2.1.1). Recall the definition of $h^{(1)}(\cdot)$ in (2.2.2). 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:

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

where $\mathcal{L}(\boldsymbol{X})$ is the distribution (law) of $\boldsymbol{X}$, and P_{jk}, $\mathrm{E}_{jk}(\cdot)$, 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 $\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathrm{P}_{jk})\}$ with regard to the dependence between X_j and X_k, characterized by the "correlation measure" $\mathrm{E}_{jk} h$. 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 2.2.1 to 2.2.3, provided γ is large enough.

Lemma 2.4.1. *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

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

Such matrices will define a "sparse local alternative" as considered also in Section 4.1 in Han et al. (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 et al. (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 et al. (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 non-consistent 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 2.4.3 (Power analysis, general)**.** *Given any $\gamma > 0$ and a kernel $h(\cdot)$ satisfying Assumption 2.2.1, there exists some sufficiently large C_γ depending on γ such that*

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

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, $U_{jk} := \mathrm{E}\widehat{U}_{jk}$.

The proof of Theorem 2.4.3 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 2.3.1–2.3.3 have power tending to 1.

Corollary 2.4.2 (Power analysis, examples). *Given any $\gamma > 0$, we have, for some sufficiently large C_γ depending on γ,*

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

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

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 $D_{jk} := \mathrm{E}\widehat{D}_{jk}$, $R_{jk} := \mathrm{E}\widehat{R}_{jk}$, and $\tau^*_{jk} := \mathrm{E}\widehat{\tau}^*_{jk}$, 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 et al. (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 et al. (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 et al. (2018) made a first step towards 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 2.3.1–2.3.3 and show their rate-optimality in the Gaussian model. To this end, we define a new alternative class of matrices

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

where $\mathbf{M} \succeq 0$ denotes positive semi-definiteness. We then have the following theorem as a consequence of Corollary 2.4.2. 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 2.4.4 (Power analysis, Gaussian)**.** *For a sufficiently large absolute constant $C_0 > 0$, we have, as long as $n, p \to \infty$,*

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

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

The proof of Theorem 2.4.4 is given in the supplement. It relies on Lemma 2.4.1 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 et al. (2017), Theorem 2.4.4 yields minimax rate-optimality of the tests in Examples 2.3.1–2.3.3 against the sparse Gaussian alternative.

Theorem 2.4.5 (Rate optimality, Theorem 5 in Han et al., 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*

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

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

2.5 Simulation studies

In this section we compare the finite-sample performance of the three tests (Extreme D, Extreme R, and Extreme τ^*) from Section 2.3 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 2.3.1;
- DHS_R: the maximum-type test in Example 2.3.2;
- DHS_{τ^*}: the maximum-type test in Example 2.3.3;
- LD_τ: the L_2-type test based on Kendall's τ (Leung and Drton, 2018);
- LD_ρ: the L_2-type test based on Spearman's ρ (Leung and Drton, 2018);
- LD_{τ^*}: the L_2-type test based on Bergsma–Dassios–Yanagimoto's τ^* (Leung and Drton, 2018);
- HCL_τ: the maximum-type test based on Kendall's τ (Han et al., 2017);
- HCL_ρ: the maximum-type test based on Spearman's ρ (Han et al., 2017);
- YZS: the L_2-type test based on the distance covariance statistic (Yao et al., 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).

2.5.1 Computational aspects

Throughout this section $\{\boldsymbol{z}_i = (z_{1i}, z_{2i})^\top\}_{i\in[\![n]\!]}$ 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 (2.3.2).

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

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

and r_i and s_i are the ranks of z_{1i} among $\{z_{11}, \ldots, z_{1n}\}$ and z_{2i} among $\{z_{21}, \ldots, z_{2n}\}$, respectively. Moreover, c_i is the number of samples $\boldsymbol{z}_{i'} = (z_{1i'}, z_{2i'})$ for which $z_{1i'} < z_{1i}$ and $z_{2i'} < z_{2i}$.

Weihs et al. (2016) and Heller and Heller (2016b) proposed algorithms for efficient computation of the Bergsma–Dassios–Yanagimoto statistic $\widehat{\tau}^*$. Without loss of generality, let $z_{11} < \cdots < z_{1n}$, i.e., $r_i = i$. Weihs et al. (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 $\ell < \ell'$,

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

Weihs et al. (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 (2016b) 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,

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

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_{2i} in $\{z_{21}, \ldots, z_{2n}\}$. As a consequence, formula (2.3.2) 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 2.1 shows empirical computation times for the considered statistics on 1,000 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`.

Table 2.1: A comparison of computation time for all the correlation statistics considered. The computation time here is the averaged elapsed time (in milliseconds) of 1,000 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

Blum–Kiefer–Rosenblatt's $\widehat{R}$ was then obtained using identity (2.3.2), 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.

2.5.2 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 $5,000$ 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 2.5.1.

(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 2.2. Those of the three distribution-dependent tests are given in Table 2.3. 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 2.2, as long as n and p grow simultaneously, a trend to the nominal level 0.05 is clear; e.g., 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 2.4.1, and this trend persists in more simulations as n and p become even larger.

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 et al. (2013) and Section 5.2 in Leung and Drton (2018). To reflect this, we consider two sets of examples

Table 2.2: Empirical sizes of the eight rank-based tests in Example 2.5.1

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

that focus on relatively sparse settings (modified based on Yao et al. (2018) and Han et al. (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 2.5.2.

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

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

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

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

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 $\boldsymbol{X}_1$.

Example 2.5.3.

(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

Table 2.3: Empirical sizes of the three distribution-dependent tests in Example 2.5.1

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

of four nonzero entries randomly from the upper triangle of $\mathbf{\Delta}$, 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 $\mathbf{\Delta}$ symmetric. Finally,

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

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

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

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

Example 2.5.4. 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 2.5.2–2.5.4 are reported in Tables 2.4–2.6. 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 2.5.2, 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 2.5.2, indicating that they are powerless in detecting the considered non-linear, non-monotone dependences, an observation that was also made in Yao et al. (2018). Fourth, Tables 2.4 and 2.5 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 2.5.3, 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 non-linear non-monotone dependences. Fifth, under a Gaussian parametric model, Table 2.5 (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 2.6 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 2.2.1, 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

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

Table 2.4: Empirical powers of the eleven competing tests in Example 2.5.2

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 2.5.2(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 2.5.2(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

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

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

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

Table 2.5: Empirical powers of the eleven competing tests in Example 2.5.3

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 2.5.3(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 2.5.3(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 2.5.3(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

Table 2.6: Empirical powers of the eleven competing tests in Example 2.5.4

n	p	DHS_D	DHS_R	DHS_{τ^*}	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 2.5.4(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 2.5.4(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 2.5.4(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

shown explicitly in the following proposition.

Proposition 2.5.1. *Under the independence hypothesis H_0, for each (n, p), we have with probability at least $1 - 2/M^2$ that*

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

Table A.1 in the supplement gives the sizes and powers of the proposed tests with simulation-based critical values ($M = 5,000$). The table shows results only for Examples 2.5.1, 2.5.3, and 2.5.4 as the simulated powers under Example 2.5.2 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.

2.6 Discussion

2.6.1 Discussion of Assumption 2.2.1

Assumption 2.2.1 plays a key role in our analysis. It synthesizes crucial properties satisfied by the three rank correlation statistics from Examples 2.2.1–2.2.3.

From a more general perspective, one might ask whether there is an exact relation between Assumption 2.2.1 and the properties of I- and D-consistency summarized in Weihs et al. (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, 2021), 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 specific and related to the particular properties of rank-based consistent tests. This assumption, however, is key to the establishment of Theorem 2.4.2.

2.6.2 Discussion of τ^*

In this section we give new perspectives on Bergsma–Dassios–Yanagimoto's correlation measure $\tau^* := \mathrm{E}h_{\tau^*}$, introduced in Example 2.2.3. 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 $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6 \in \mathbb{R}^2$ have no tie among their first and their second entries, respectively, then

$$3 \cdot \binom{6}{5}^{-1} \sum_{1 \le i_1 < \cdots < i_5 \le 6} h_D(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_5}) + 2h_R(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6) \tag{2.6.1}$$
$$= 5 \cdot \binom{6}{4}^{-1} \sum_{1 \le i_1 < \cdots < i_4 \le 6} h_{\tau^*}(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_4}).$$

Equation (2.6.1) 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 $\mathrm{E}h_{\tau^*} \ge 0$ with equality if and only if Z_1 and Z_2 are independent?

Theorem 2.6.1. *For any random vector $\boldsymbol{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$ with continuous marginal distributions, we have $\mathrm{E}h_{\tau^*} \ge 0$ and the equality holds if and only if Z_1 is independent of Z_2.*

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

Theorem 2.6.2. *If $\boldsymbol{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$ is bivariate Gaussian with (Pearson) correlation ρ, then $\mathrm{E}h_D$ and $\mathrm{E}h_R$ and, thus, also $\mathrm{E}h_{\tau^*}$ are increasing functions of $|\rho|$.*

Theorem 2.6.1 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 et al. (2018) who noted that based on existing literature "it is not guaranteed that $\mathrm{E}h_{\tau^*} > 0$ 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 2.6.1. *For $(X, Y)^\top$ following the uniform distribution on the unit circle in $\mathbb{R}^2$, we have $\mathrm{E}h_D = \mathrm{E}h_R = \mathrm{E}h_{\tau^*} = 1/16$.*

Chapter 3

ON UNIVERSALLY CONSISTENT AND FULLY DISTRIBUTION-FREE RANK TESTS OF VECTOR INDEPENDENCE

3.1 Introduction

Quantifying the dependence between two variables and testing for their independence are among the oldest and most fundamental problems of statistical inference. The (marginal) distributions of the two variables under study, in that context, typically play the role of nuisances, and the need for a nonparametric approach naturally leads, when they are univariate, to distribution-free methods based on their ranks. This paper is dealing with the multivariate extension of that approach.

3.1.1 Measuring vector dependence and testing independence

Consider two absolutely continuous random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. The problems of measuring the dependence between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ and testing their independence when $d_1 = d_2 = 1$ (call this the univariate case) have a long history that goes back more than a century (Pearson, 1895; Spearman, 1904). The same problem when d_1 and d_2 are possibly unequal and larger than one (the multivariate case) is of equal practical interest but considerably more challenging. Following early attempts (Wilks, 1935), a large literature has emerged, with renewed interest in recent years.

When the marginal distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are unspecified and $d_1 = d_2 = 1$, rank correlations provide a natural and appealing nonparametric approach to testing for independence, as initiated in the work of Spearman (1904) and Kendall (1938); cf. Chapter III.6

in Hájek and Šidák (1967). On one hand, ranks yield distribution-free tests because, under the null hypothesis of independence, their distributions do not depend on the unspecified marginal distributions. On the other hand, they can be designed (Hoeffding, 1948; Blum et al., 1961; Bergsma and Dassios, 2014; Yanagimoto, 1970) to consistently estimate dependence measures that vanish if and only if independence holds, and so detect any type of dependence—something Spearman and Kendall's rank correlations cannot.

New subtleties arise, however, when attempting to extend the rank-based approach to the multivariate case. While d_k ranks can be constructed separately for each coordinate of $\boldsymbol{X}_k$, $k = 1, 2$, their joint distribution depends on the distribution of the underlying $\boldsymbol{X}_k$, preventing distribution-freeness of the (d_1+d_2)-tuple of ranks. As a consequence, the existing tests of multivariate independence based on componentwise ranks (e.g., Puri et al., 1970) are not distribution-free, which has both computational implications (e.g., through a need for permutation analysis) and statistical implications (as we shall detail soon).

3.1.2 Desirable properties

In this paper, we develop a general framework for multivariate analogues of popular rank-based measures of dependence for the univariate case. Our objective is to achieve the following five desirable properties.

(1) *Full distribution-freeness.* Many statistical tests exploit asymptotic distribution-freeness for computationally efficient distributional approximations yielding pointwise asymptotic control of their size. This is the case, for instance, with Hallin and Paindaveine (2002c,b,a, 2008) due to estimation of a scatter matrix, or with Taskinen et al. (2003, 2004), Taskinen et al. (2005). Pointwise asymptotics yield, for any given significance level $\alpha \in (0, 1)$, a sequence of tests $\phi_\alpha^{(n)}$ indexed by the sample size n such that $\lim_{n\to\infty} \mathrm{E}_{\mathrm{P}}[\phi_\alpha^{(n)}] = \alpha$ for every distribution P from a class $\mathcal{P}$ of null distributions. Generally, however, the size fails to be controlled in a uniform sense, that is, it does not hold that $\lim_{n\to\infty} \sup_{\mathrm{P}\in\mathcal{P}} \mathrm{E}_{\mathrm{P}}[\phi_\alpha^{(n)}] \le \alpha$,

which may explain poor finite-sample properties (see, e.g., Le Cam and Yang, 2000; Leeb and Pötscher, 2008; Belloni et al., 2014). While uniform inferential validity is impossible to achieve for some problems, e.g., when testing for conditional independence (Shah and Peters, 2020; Azadkia and Chatterjee, 2021), we shall see that it is achievable for testing (unconditional) multivariate independence. Indeed, for fully distribution-free tests, as obtained from our rank-based approach, pointwise validity automatically implies uniform validity.

(2) *Transformation invariance.* A dependence measure μ is said to be invariant under orthogonal transformations, shifts, and global rescaling if

$$\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu(\boldsymbol{v}_1 + a_1\mathbf{O}_1\boldsymbol{X}_1, \boldsymbol{v}_2 + a_2\mathbf{O}_1\boldsymbol{X}_2)$$

for any scalars $a_k > 0$, vectors $\boldsymbol{v}_k \in \mathbb{R}^{d_k}$, and orthogonal $d_k \times d_k$ matrices $\mathbf{O}_k$, $k = 1, 2$. This invariance, here simply termed "transformation invariance", is a natural requirement in cases where the components of $\boldsymbol{X}_1, \boldsymbol{X}_2$ do not have specific meanings and observations could have been recorded in another coordinate system. Such invariance is of considerable interest in multivariate statistics (see, e.g., Gieser and Randles, 1997; Taskinen et al., 2003, 2005; Oja et al., 2016).

(3) *Consistency.* Weihs et al. (2018) call a dependence measure μ *I-consistent* within a family of distributions $\mathcal{P}$ if independence between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with joint distribution in $\mathcal{P}$ implies $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$. If $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$ implies independence of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ (i.e., dependence of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ implies $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) \neq 0$), then μ is *D-consistent* within $\mathcal{P}$. Note that the measures considered in this paper do not necessarily take maximal value 1 if and only if one random vector is a measurable function of the other. While any reasonable dependence measure should be I-consistent, prominent examples (Pearson's correlation, Spearman's ρ, Kendall's τ) fail to be D-consistent. If a dependence measure μ is I- and D-consistent, then the consistency of tests based on an estimator $\mu^{(n)}$ of μ is guaranteed by the (strong or weak) consistency of that estimator. Dependence measures that are both I- and D-consistent

(within a large nonparametric family) serve an important purpose as they are able to capture nonlinear dependences. Well-known I- and D-consistent measures for the univariate case include Hoeffding's D (Hoeffding, 1948), Blum–Kiefer–Rosenblatt's R (Blum et al., 1961), and Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios, 2014; Yanagimoto, 1970; Drton et al., 2020). Multivariate extensions have been proposed, e.g., in Gretton et al. (2005c), Székely et al. (2007), Heller et al. (2012), Heller et al. (2013), Heller and Heller (2016a), Zhu et al. (2017), Weihs et al. (2018), Kim et al. (2020b), Deb and Sen (2021), Shi et al. (2021a), Berrett et al. (2021).

(4) *Statistical efficiency.* Once its size is controlled, the performance of a test may be evaluated through its power against local alternatives. For the proposed tests, our focus is on quadratic mean differentiable alternatives (Lehmann and Romano, 2005, Sec. 12.2), which form a popular class for conducting local power analyses; for related recent examples see Bhattacharya (2019, Section 3) and Cao and Bickel (2020, Section 4.4). Our results then show the nontrivial local power of our tests in $n^{-1/2}$ neighborhoods within this class.

(5) *Computational efficiency.* Statistical properties aside, modern applications require the evaluation of a dependence measure and the corresponding test to be as computationally efficient as possible. We thus prioritize measures leading to low computational complexity.

The main challenge, with this list of five properties, lies in combining the full distribution-freeness from property (1) with properties (2)–(5). The solution, as we shall see, involves an adequate multivariate extension of the univariate concepts of ranks and signs.

3.1.3 Contribution of this paper

This paper proposes a class of dependence measures and tests that achieve the five properties from Section 3.1.2 by leveraging the recently introduced multivariate center-outward ranks and signs (Chernozhukov et al., 2017; Hallin, 2017); see Hallin et al. (2021a) for a complete account. In contrast to earlier related concepts such as componentwise ranks (Puri and Sen, 1971), spatial ranks (Oja, 2010; Han and Liu, 2018), depth-based ranks (Liu and Singh, 1993; Zuo and He, 2006), and pseudo-Mahalanobis ranks and signs (Hallin and Paindaveine, 2002c), the new concept yields statistics that enjoy full distribution-freeness (in finite samples and, thus, asymptotically) as soon as the underlying probability measure is Lebesgue-absolutely continuous. This allows for a general multivariate strategy, in which the observations are replaced by functions of their center-outward ranks and signs when forming dependence measures and corresponding test statistics. This is also the idea put forth in Shi et al. (2021a) and, in a slightly different way, in Deb and Sen (2021), where the focus is on distance covariance between center-outward ranks and signs.

Methodologically, we are generalizing this approach in two important ways. First, we introduce a class of *generalized symmetric covariances* (GSCs) along with their center-outward rank versions, of which the distance covariance concepts from Deb and Sen (2021) and Shi et al. (2021a) are but particular cases. Second, we show how considerable additional flexibility and power results from incorporating score functions in the definition. Our simulations in Section 3.5.4 exemplify the benefits of this "score-based" approach.

From a theoretical point of view, we offer a new approach to asymptotic theory for the proposed rank-based statistics. Indeed, handling this general class with the methods of Shi et al. (2021a) or Deb and Sen (2021) would be highly nontrivial. Moreover, these methods would not provide any insights into local power—an issue receiving much attention also in other contexts (Hallin et al., 2021b; Beirlant et al., 2020; Hallin et al., 2021c, 2020). We thus develop a completely different method, based on a general asymptotic representation result applicable to all center-outward rank-based GSCs under the null hypothesis of independence and contiguous alternatives of dependence. Our result (Theorem 3.5.1) is a multivariate extension of Hájek's classical asymptotic representation for univariate linear rank statistics (Hájek and Šidák, 1967) and also simplifies the derivation of limiting null distributions.

Combined with a nontrivial use of Le Cam's third lemma in a context of non-Gaussian limits, our approach allows for the first local power results in the area; the statistical efficiency of the tests of Deb and Sen (2021) and Shi et al. (2021a) follows as a special case. In Proposition 3.4.2, we establish the strong consistency of our rank-based tests against any fixed alternative under a regularity condition on the score function. Thanks to a recent result by Deb et al. (2021), that assumption can be relaxed: our tests, thus, enjoy *universal consistency* against fixed dependence alternatives.

Outline of the paper The paper begins with a review of important dependence measures from the literature (Section 3.2). Generalizing the idea of symmetric rank covariances put forth in Weihs et al. (2018), we show that a single formula unifies them all; we term the concept *generalized symmetric covariance* (GSC). As further background, Section 3.3 introduces the notion of center-outward ranks and signs. Section 3.4 presents our streamlined approach of defining multivariate dependence measures, along with sample counterparts, and highlights some of their basic properties. Section 3.5 treats tests of independence and develops a theory of asymptotic representation for center-outward rank-based GSCs (Section 3.5.1) as well as the local power analysis of the corresponding tests against classes of quadratic mean differentiable alternatives (Section 3.5.2). Specific alternatives are exemplified in Section 3.5.3, and benefits of choosing standard score functions (such as normal scores) are illustrated in the numerical study in Section 3.5.4. All proofs are deferred to the appendix.

Notation For integer $m \geq 1$, put $[\![m]\!] := \{1, 2, \ldots, m\}$, and let $\mathfrak{S}_m$ be the symmetric group, i.e., the group of all permutations of $[\![m]\!]$. We write $\mathrm{sgn}(\sigma)$ for the sign of $\sigma \in \mathfrak{S}_m$. In the sequel, the subgroup

$$H_*^m := \langle (1\ 4), (2\ 3) \rangle = \{(1), (1\ 4), (2\ 3), (1\ 4)(2\ 3)\} \subset \mathfrak{S}_m \tag{3.1.1}$$

will play an important role. Here, we have made use of the cycle notation (omitting 1-cycles) so that, e.g., (1) denotes the identity permutation and

$$(1\ 4) \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & m \\ 4 & 2 & 3 & 1 & 5 & 6 & \cdots & m \end{pmatrix}, \qquad (1\ 4)(2\ 3) \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & m \\ 4 & 3 & 2 & 1 & 5 & 6 & \cdots & m \end{pmatrix},$$

where the right-hand sides are in classical two-line notation listing $\sigma(i)$ below i, $i \in [\![m]\!]$.

A set with distinct elements $x_1, \ldots, x_n$ is written either as $\{x_1, \ldots, x_n\}$ or $\{x_i\}_{i=1}^n$. The corresponding sequence is denoted by $[x_1, \ldots, x_n]$ or $[x_i]_{i=1}^n$. An arrangement of $\{x_i\}_{i=1}^n$ is a sequence $[x_{\sigma(i)}]_{i=1}^n$, where $\sigma \in \mathfrak{S}_n$. An r-arrangement is a sequence $[x_{\sigma(i)}]_{i=1}^r$ for $r \in [\![n]\!]$. Write I_r^n for the family of all $(n)_r := n!/(n-r)!$ possible r-arrangements of $[\![n]\!]$.

The set of nonnegative reals is denoted $\mathbb{R}_{\geq 0}$, and $\mathbf{0}_d$ stands for the origin in $\mathbb{R}^d$. For two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$, we write $\boldsymbol{u} \preceq \boldsymbol{v}$ if $u_\ell \leq v_\ell$ for all $\ell \in [\![d]\!]$, and $\boldsymbol{u} \not\preceq \boldsymbol{v}$ otherwise. Let $\mathsf{Arc}(\boldsymbol{u}, \boldsymbol{v}) := (2\pi)^{-1} \arccos\{\boldsymbol{u}^\top \boldsymbol{v}/(\|\boldsymbol{u}\|\|\boldsymbol{v}\|)\}$ if $\boldsymbol{u}, \boldsymbol{v} \neq \mathbf{0}_d$; $\mathsf{Arc}(\boldsymbol{u}, \boldsymbol{v}) := 0$ otherwise. Here, $\|\cdot\|$ stands for the Euclidean norm. For vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, we use $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ as a shorthand for $(\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_k^\top)^\top$. We write $\mathbf{I}_d$ for the $d \times d$ identity matrix. For a function $f : \mathcal{X} \to \mathbb{R}$, we define $\|f\|_\infty := \max_{x \in \mathcal{X}} |f(x)|$. The symbols $\lfloor \cdot \rfloor$ and $\mathbb{1}(\cdot)$ stand for the floor and indicator functions.

The cumulative distribution function and the probability distribution of a real-valued random variable/vector $\boldsymbol{Z}$ are denoted as $F_{\boldsymbol{Z}}(\cdot)$ and $\mathrm{P}_{\boldsymbol{Z}}$, respectively. The class of probability measures on $\mathbb{R}^d$ that are absolutely continuous (with respect to the Lebesgue measure) is denoted as $\mathcal{P}_d^{\mathrm{ac}}$. We use $\rightsquigarrow$ and $\xrightarrow{\text{a.s.}}$ to denote convergence in distribution and almost sure convergence, respectively. For any symmetric kernel $h(\cdot)$ on $(\mathbb{R}^d)^m$, any integer $\ell \in [\![m]\!]$, and any probability measure $\mathrm{P}_{\boldsymbol{Z}}$, we write $h_\ell(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}})$ for $\mathrm{E}h(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell, \boldsymbol{Z}_{\ell+1}, \ldots, \boldsymbol{Z}_m)$ where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are m independent copies of $\boldsymbol{Z} \sim \mathrm{P}_{\boldsymbol{Z}}$, and $\mathrm{E}h := \mathrm{E}h(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m)$. The product measure of two distributions P_1 and P_2 is denoted $\mathrm{P}_1 \otimes \mathrm{P}_2$.

3.2 Generalized symmetric covariances

Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be two random vectors with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively, and assume throughout this paper that they are both absolutely continuous with respect to the Lebesgue measure. Weihs et al. (2018, Def. 3) introduced a general approach to defining rank-based measures of dependence via signed sums of indicator functions that are acted upon by subgroups of the symmetric group. In this section, we highlight that their resulting family of *symmetric rank covariances* can be extended to cover a much wider range of dependence measures including, in particular, the celebrated *distance covariance* (Székely et al., 2007). This enables us to handle a broad family of dependence measures in the following common standard form.

Definition 3.2.1 (Generalized symmetric covariance)**.** A measure of dependence μ is said to be an m-th order *generalized symmetric covariance (GSC)* if there exist two kernel functions $f_1 : (\mathbb{R}^{d_1})^m \to \mathbb{R}_{\geq 0}$ and $f_2 : (\mathbb{R}^{d_2})^m \to \mathbb{R}_{\geq 0}$, and a subgroup $H \subseteq \mathfrak{S}_m$ containing an equal number of even and odd permutations such that

$$\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_{f_1,f_2,H}(\boldsymbol{X}_1, \boldsymbol{X}_2) := \mathrm{E}[k_{f_1,f_2,H}((\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1m}, \boldsymbol{X}_{2m}))].$$

Here $(\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1m}, \boldsymbol{X}_{2m})$ are m independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, and the dependence kernel function $k_{f_1,f_2,H}(\cdot)$ is defined as

$$k_{f_1,f_2,H}\Big((\boldsymbol{x}_{11}, \boldsymbol{x}_{21}), \ldots, (\boldsymbol{x}_{1m}, \boldsymbol{x}_{2m})\Big) := \Big\{\sum_{\sigma \in H} \mathrm{sgn}(\sigma) f_1(\boldsymbol{x}_{1\sigma(1)}, \ldots, \boldsymbol{x}_{1\sigma(m)})\Big\}\Big\{\sum_{\sigma \in H} \mathrm{sgn}(\sigma) f_2(\boldsymbol{x}_{2\sigma(1)}, \ldots, \boldsymbol{x}_{2\sigma(m)})\Big\}. \tag{3.2.1}$$

As the group H is required to have equal numbers of even and odd permutations, the order of a GSC satisfies $m \geq 2$. This requirement also justifies the term "generalized covariance" through the following property; compare Weihs et al. (2018, Prop. 2).

Proposition 3.2.1. *All GSCs are I-consistent. More precisely, the GSC* $\mu_{f_1,f_2,H}(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *is I-consistent in the family of distributions such that* $\mathrm{E}[f_k] := \mathrm{E}[f_k(\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km})] < \infty$, $k = 1, 2$, *where* $\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km}$ *are* m *independent copies of* $\boldsymbol{X}_k$.

The concept of GSC unifies a surprisingly large number of well-known dependence measures. We consider here five noteworthy examples, namely, the distance covariance of Székely et al. (2007) and Székely and Rizzo (2013), the multivariate version of Hoeffding's D based on marginal ordering (Weihs et al., 2018, Section 2.2, p. 549), and the projection-averaging extensions of Hoeffding's D (Zhu et al., 2017), of Blum–Kiefer–Rosenblatt's R (Kim et al., 2020c, Proposition D.5), and of Bergsma–Dassios–Yanagimoto's τ^* (Kim et al., 2020b, Theorem. 7.2). Only one type of subgroup, namely, $H_*^m := \langle (1\ 4), (2\ 3) \rangle \subseteq \mathfrak{S}_m$ for $m \geq 4$ is needed; recall (3.1.1). For simplicity, we write $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) \mapsto f_k(\boldsymbol{w})$ for the kernel functions of an mth order multivariate GSC for which the dimension of $\boldsymbol{w}_\ell$, $\ell = 1, \ldots, m$, is d_k, hence may differ for $k = 1$ and $k = 2$. Not all components of $\boldsymbol{w}$ need to have an impact on $f_k(\boldsymbol{w})$. For instance, the kernels of distance covariance, a 4th order GSC, map $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_4)$ to $\mathbb{R}_{\geq 0}$ but depend neither on $\boldsymbol{w}_3$ nor $\boldsymbol{w}_4$.

Example 3.2.1 (Examples of multivariate GSCs)**.**

(a) Distance covariance is a 4th order GSC with $H = H_*^4$ and

$$f_k^{\mathrm{dCov}}(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}_1 - \boldsymbol{w}_2\| \quad \text{on} \quad (\mathbb{R}^{d_k})^4, \quad k = 1, 2.$$

Indeed, with $c_d := \pi^{(1+d)/2}/\Gamma((1+d)/2)$, we have

$$\begin{aligned} &\mu_{f_1^{\mathrm{dCov}}, f_2^{\mathrm{dCov}}, H_*^4}(\boldsymbol{X}_1, \boldsymbol{X}_2) \\ &\quad = \frac{1}{4}\mathrm{E}[(\|\boldsymbol{X}_{11} - \boldsymbol{X}_{12}\| - \|\boldsymbol{X}_{11} - \boldsymbol{X}_{13}\| - \|\boldsymbol{X}_{14} - \boldsymbol{X}_{12}\| + \|\boldsymbol{X}_{14} - \boldsymbol{X}_{13}\|) \\ &\qquad\qquad \times (\|\boldsymbol{X}_{21} - \boldsymbol{X}_{22}\| - \|\boldsymbol{X}_{21} - \boldsymbol{X}_{23}\| - \|\boldsymbol{X}_{24} - \boldsymbol{X}_{22}\| + \|\boldsymbol{X}_{24} - \boldsymbol{X}_{23}\|)] \\ &\quad = \frac{1}{c_{d_1} c_{d_2}} \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} \frac{|\varphi_{(\boldsymbol{X}_1, \boldsymbol{X}_2)}(\boldsymbol{t}_1, \boldsymbol{t}_2) - \varphi_{\boldsymbol{X}_1}(\boldsymbol{t}_1)\varphi_{\boldsymbol{X}_2}(\boldsymbol{t}_2)|^2}{\|\boldsymbol{t}_1\|^{d_1+1}\|\boldsymbol{t}_2\|^{d_2+1}} \mathrm{d}\boldsymbol{t}_1 \mathrm{d}\boldsymbol{t}_2. \end{aligned} \tag{3.2.2}$$

Identity (3.2.2) was established in Székely et al. (2007, Remark 3), Székely and Rizzo (2009, Thm. 8), and Bergsma and Dassios (2014, Sec. 3.4);

(b) Hoeffding's multivariate marginal ordering D is a 5th order GSC with $H = H_*^5$ and

$$f_k^M(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\boldsymbol{w}_1, \boldsymbol{w}_2 \preceq \boldsymbol{w}_5) \;\text{ on }\; (\mathbb{R}^{d_k})^5, \quad k = 1, 2,$$

since, by Weihs et al. (2018, Prop. 1),

$$\mu_{f_1^M, f_2^M, H_*^5}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} \{F_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}(\boldsymbol{u}_1, \boldsymbol{u}_2) - F_{\boldsymbol{X}_1}(\boldsymbol{u}_1)F_{\boldsymbol{X}_2}(\boldsymbol{u}_2)\}^2 \mathrm{d}F_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}(\boldsymbol{u}_1, \boldsymbol{u}_2);$$

(c) Hoeffding's multivariate projection-averaging D is a 5th order GSC with $H = H_*^5$ and

$$f_k^D(\boldsymbol{w}) = \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_5, \boldsymbol{w}_2 - \boldsymbol{w}_5) \;\text{ on }\; (\mathbb{R}^{d_k})^5, \quad k = 1, 2.$$

Indeed, by Zhu et al. (2017, Equation (3)), we have

$$\begin{aligned}\mu_{f_1^D, f_2^D, H_*^5}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathcal{S}_{d_1-1}\times\mathcal{S}_{d_2-1}} \int_{\mathbb{R}^2} \{&F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2) \\ &- F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1) F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\}^2 \mathrm{d}F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2) \mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1) \mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2),\end{aligned}$$

with λ_d the uniform measure on the unit sphere $\mathcal{S}_{d-1}$;

(d) Blum–Kiefer–Rosenblatt's multivariate projection-averaging R is a 6th order GSC with $H = H_*^6$ and

$$\begin{aligned} f_1^R(\boldsymbol{w}) &= \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_5, \boldsymbol{w}_2 - \boldsymbol{w}_5) \;\text{ on }\; (\mathbb{R}^{d_1})^6, \\ f_2^R(\boldsymbol{w}) &= \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_6, \boldsymbol{w}_2 - \boldsymbol{w}_6) \;\text{ on }\; (\mathbb{R}^{d_2})^6; \end{aligned}$$

this follows from Kim et al. (2020c, Prop. D.5), who showed

$$\mu_{f_1^R, f_2^R, H_*^6}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathcal{S}_{d_1-1}\times\mathcal{S}_{d_2-1}} \int_{\mathbb{R}^2} \{F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2)$$

$$-F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1)F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\}^2 \mathrm{d}F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1)\mathrm{d}F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1)\mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2);$$

(e) Bergsma–Dassios–Yanagimoto's multivariate projection-averaging τ^* is a 4th order GSC with $H = H_*^4$ and

$$f_k^{\tau^*}(\boldsymbol{w}) = \mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_2, \boldsymbol{w}_2 - \boldsymbol{w}_3) + \mathsf{Arc}(\boldsymbol{w}_2 - \boldsymbol{w}_1, \boldsymbol{w}_1 - \boldsymbol{w}_4) \quad \text{on} \quad (\mathbb{R}^{d_k})^4, \quad k = 1, 2,$$

since, by Kim et al. (2020b, Theorem 7.2), we have

$$\mu_{f_1^{\tau^*}, f_2^{\tau^*}, H_*^4}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathcal{S}_{d_1-1}\times\mathcal{S}_{d_2-1}} \mathrm{E}\{a_{\mathrm{sign}}(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_{11}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{12}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{13}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{14})$$
$$\times\, a_{\mathrm{sign}}(\boldsymbol{\alpha}_2^\top \boldsymbol{X}_{21}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{22}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{23}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{24})\}\mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1)\mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2),$$

with $a_{\mathrm{sign}}(w_1, w_2, w_3, w_4) := \mathrm{sign}(|w_1 - w_2| - |w_1 - w_3| - |w_4 - w_2| + |w_4 - w_3|)$.

Remark 3.2.1. Sejdinovic et al. (2013) recognize distance covariance as an example of an HSIC-type statistic (Gretton et al., 2005c,a,b; Fukumizu et al., 2007). The HSIC-type statistics are all 4th order multivariate GSCs, and we note that our results for distance covariance readily extend to other HSIC-type statistics.

Remark 3.2.2. In the univariate case, the GSCs from Example 3.2.1(b)–(e) reduce to the D of Hoeffding (1948), R of Blum et al. (1961), and τ^* of Bergsma and Dassios (2014), respectively. As shown by Drton et al. (2020), the latter is connected to the work of Yanagimoto (1970). In Appendix B.2.1, we simplify the kernels for the univariate case, and show that the GSC framework also covers the τ of Kendall (1938).

All the multivariate dependence measures we have introduced are D-consistent, albeit with some variations in the families of distributions for which this holds; see, e.g., the discussions in Examples 2.1–2.3 of Drton et al. (2020). As these dependence measures all involve the group H_*^m, we highlight the following fact.

Lemma 3.2.1. *A GSC $\mu = \mu_{f_1,f_2,H_*^m}$ with $m \geq 4$ is D-consistent in a family $\mathcal{P}$ if and only if the pair (f_1, f_2) is D-consistent in $\mathcal{P}$—namely, if and only if*

$$\mathrm{E}\Big[\prod_{k=1}^{2}\Big\{f_k(\boldsymbol{X}_{k1},\boldsymbol{X}_{k2},\boldsymbol{X}_{k3},\boldsymbol{X}_{k4},\boldsymbol{X}_{k5},\ldots,\boldsymbol{X}_{km}) - f_k(\boldsymbol{X}_{k1},\boldsymbol{X}_{k3},\boldsymbol{X}_{k2},\boldsymbol{X}_{k4},\boldsymbol{X}_{k5},\ldots,\boldsymbol{X}_{km}) \\ - f_k(\boldsymbol{X}_{k4},\boldsymbol{X}_{k2},\boldsymbol{X}_{k3},\boldsymbol{X}_{k1},\boldsymbol{X}_{k5},\ldots,\boldsymbol{X}_{km}) + f_k(\boldsymbol{X}_{k4},\boldsymbol{X}_{k3},\boldsymbol{X}_{k2},\boldsymbol{X}_{k1},\boldsymbol{X}_{k5},\ldots,\boldsymbol{X}_{km})\Big\}\Big]$$

is finite, nonnegative, and equal to 0 only if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent.

Theorem 3.2.1. *All the multivariate GSCs in Example 3.2.1 are D-consistent within the family $\big\{\mathrm{P} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2} \big| \mathrm{E}_{\mathrm{P}}[f_k(\boldsymbol{X}_{k1},\ldots,\boldsymbol{X}_{km})] < \infty,\ k = 1,2\big\}$ (with f_k, $k = 1,2$ denoting their respective kernels).*

The invariance/equivariance properties of GSCs depend on those of their kernels. We say that a kernel function $f : (\mathbb{R}^d)^m \to \mathbb{R}$ is *orthogonally invariant* if, for any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{d\times d}$ and any $\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m \in (\mathbb{R}^d)^m$, $f(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m) = f(\mathbf{O}\boldsymbol{w}_1,\ldots,\mathbf{O}\boldsymbol{w}_m)$.

Lemma 3.2.2. *If f_1 and f_2 both are orthogonally invariant, then any GSC of the form $\mu = \mu_{f_1,f_2,H}$ is orthogonally invariant, i.e., $\mu(\boldsymbol{X}_1,\boldsymbol{X}_2) = \mu(\mathbf{O}_1\boldsymbol{X}_1,\mathbf{O}_2\boldsymbol{X}_2)$ for any pair of random vectors $(\boldsymbol{X}_1,\boldsymbol{X}_2)$ and orthogonal matrices $\mathbf{O}_1 \in \mathbb{R}^{d_1\times d_1}$ and $\mathbf{O}_2 \in \mathbb{R}^{d_2\times d_2}$.*

Proposition 3.2.2. *The kernels (a),(c)–(e) in Example 3.2.1, hence the corresponding GSCs, are orthogonally invariant.*

Turning from theoretical dependence measures to their empirical counterparts, it is clear that any GSC admits a natural unbiased estimator in the form of a U-statistic, which we call the *sample generalized symmetric covariance* (SGSC).

Definition 3.2.2 (Sample generalized symmetric covariance)**.** The sample generalized symmetric covariance of $\mu = \mu_{f_1,f_2,H}$ is $\widehat{\mu}^{(n)} = \widehat{\mu}^{(n)}([(\boldsymbol{x}_{1i},\boldsymbol{x}_{2i})]_{i=1}^n; f_1, f_2, H)$, of the form

$$\widehat{\mu}^{(n)} = \binom{n}{m}^{-1} \sum_{i_1<i_2<\cdots<i_m} \overline{k}_{f_1,f_2,H}\Big((\boldsymbol{x}_{1i_1},\boldsymbol{x}_{2i_1}),\ldots,(\boldsymbol{x}_{1i_m},\boldsymbol{x}_{2i_m})\Big),$$

where $\overline{k}_{f_2,f_2,H}$ is the "symmetrized" version of $k_{f_2,f_2,H}$:

$$\overline{k}_{f_1,f_2,H}\Big(\big[(\boldsymbol{x}_{1\ell},\boldsymbol{x}_{2\ell})\big]_{\ell=1}^{m}\Big) := \frac{1}{m!}\sum_{\sigma\in\mathfrak{S}_m} k_{f_1,f_2,H}\Big(\big[(\boldsymbol{x}_{1\sigma(\ell)},\boldsymbol{x}_{2\sigma(\ell)}\big]_{\ell=1}^{m}\Big).$$

If the kernels f_1 and f_2 are orthogonally invariant, then it also holds that all SGSCs of the form $\widehat{\mu}^{(n)}(\,\cdot\,; f_1, f_2, H)$ are orthogonally invariant, in the sense of remaining unaffected when the input $[(\boldsymbol{x}_{1i},\boldsymbol{x}_{2i})]_{i=1}^{n}$ is transformed into $[(\mathbf{O}_1\boldsymbol{x}_{1i},\mathbf{O}_2\boldsymbol{x}_{2i})]_{i=1}^{n}$ where $\mathbf{O}_1 \in \mathbb{R}^{d_1\times d_1}$ and $\mathbf{O}_2 \in \mathbb{R}^{d_2\times d_2}$ are arbitrary orthogonal matrices. Proposition 3.2.2 thus also implies the orthogonal invariance of SGSCs associated with kernels (a) and (c)–(e) in Example 3.2.1.

The SGSCs associated with the examples listed in Example 3.2.1, unfortunately, all fail to satisfy the crucial property of distribution-freeness. However, as we will show in Section 3.4, distribution-freeness, along with transformation invariance, can be obtained by computing SGSCs from (functions of) the center-outward ranks and signs of the observations.

3.3 Center-outward ranks and signs

This section briefly introduces the concepts of center-outward ranks and signs to be used in the sequel. The main purpose is to fix notation and terminology; for a comprehensive coverage, we refer to Hallin et al. (2021a).

We are concerned with defining multivariate ranks for a sample of d-dimensional observations drawn from a distribution in the class $\mathcal{P}_d^{\mathrm{ac}}$ of absolutely continuous probability measures on $\mathbb{R}^d$ with $d \geq 2$. Let $\mathbb{S}_d$ and $\mathcal{S}_{d-1}$ denote the open unit ball and the unit sphere in $\mathbb{R}^d$, respectively. Denote by U_d the spherical uniform measure on $\mathbb{S}_d$, that is, the product of the uniform measures on $[0,1)$ (for the distance to the origin) and on $\mathcal{S}_{d-1}$ (for the direction). The push-forward of a measure Q by a measurable transformation T is denoted as $T\sharp\mathrm{Q}$.

Definition 3.3.1 (Center-outward distribution function)**.** The *center-outward distribution function* of a probability measure $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ is the P-a.s. unique function $\mathbf{F}_\pm$ that (i) maps $\mathbb{R}^d$

to the open unit ball $\mathbb{S}_d$, (ii) is the gradient of a convex function on $\mathbb{R}^d$, and (iii) pushes P forward to U_d (i.e., such that $\mathbf{F}_\pm \sharp \mathrm{P} = \mathrm{U}_d$).

The center-outward distribution function $\mathbf{F}_\pm$ of P entirely characterizes P provided that $\mathrm{P} \in \mathcal{P}_{\mathrm{d}}^{\mathrm{ac}}$; cf. Hallin et al. (2021a, Prop. 2.1(iii)). Also, $\mathbf{F}_\pm$ is invariant under shift, global rescaling, and orthogonal transformations. We refer the readers to Appendix B.2.2 for details about these elementary properties of center-outward distribution functions.

The sample counterpart $\mathbf{F}_\pm^{(n)}$ of $\mathbf{F}_\pm$ is based on an n-tuple of data points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n \in \mathbb{R}^d$. The key idea is to construct n grid points in the unit ball $\mathbb{S}_d$ such that the corresponding discrete uniform distribution converges weakly to U_d as $n \to \infty$. For $d \geq 2$, the construction proposed in Hallin (2017, Sec. 4.2) starts by factorizing n into

$$n = n_R n_S + n_0, \qquad n_R, n_S \in \mathbb{Z}_+, \qquad 0 \leq n_0 < \min\{n_R, n_S\},$$

where in asymptotic scenarios n_R and $n_S \to \infty$, hence $n_0/n \to 0$, as $n \to \infty$. Next consider the intersection points between

– the n_R hyperspheres centered at $\mathbf{0}_d$, with radii $r/(n_R + 1)$, $r \in [\![n_R]\!]$, and

– n_S rays given by distinct unit vectors $\{\boldsymbol{s}_s^{(n_S)}\}_{s \in [\![n_S]\!]}$ that divide the unit circle into arcs of equal length $2\pi/n_S$ for $d = 2$, and are distributed as regularly as possible on the unit sphere $\mathcal{S}_{d-1}$ for $d \geq 3$; asymptotic statements merely require that the discrete uniform distribution over $\{\boldsymbol{s}_s^{(n_S)}\}_{s=1}^{n_S}$ converges weakly to the uniform distribution on $\mathcal{S}_{d-1}$ as $n_S \to \infty$.

Letting $\boldsymbol{n} := (n_R, n_S, n_0)$, the grid $\mathfrak{G}_{\boldsymbol{n}}^d$ is defined as the set of $n_R n_S$ points $\left\{\frac{r}{n_R+1}\boldsymbol{s}_s^{(n_S)}\right\}$ with $r \in [\![n_R]\!]$ and $s \in [\![n_S]\!]$ as described above along with the origin $\mathbf{0}$ in case $n_0 = 1$ or, whenever $n_0 > 1$, the n_0 points $\left\{\frac{1}{2(n_R+1)}\boldsymbol{s}_s^{(n_S)}\right\}$, $s \in \mathcal{S}$ where $\mathcal{S}$ is chosen as a random sample of size n_0 without replacement from $[\![n_S]\!]$. For $d = 1$, letting $n_S = 2$, $n_R = \lfloor n/n_S \rfloor$,

$n_0 = n - n_R n_S = 0$ or 1, $\mathfrak{G}^d_{\boldsymbol{n}}$ reduces to the points $\{\pm r/(n_R+1) : r \in [\![n_R]\!]\}$, along with the origin $\mathbf{0}$ in case $n_0 = 1$.

The empirical version $\mathbf{F}^{(n)}_{\pm}$ of $\mathbf{F}_{\pm}$ is then defined as the optimal coupling between the observed data points and the grid $\mathfrak{G}^d_{\boldsymbol{n}}$.

Definition 3.3.2 (Center-outward ranks and signs)**.** Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be distinct data points in $\mathbb{R}^d$. Let $\mathcal{T}$ be the collection of all bijective mappings between the set $\{\boldsymbol{z}_i\}_{i=1}^n$ and the grid $\mathfrak{G}^d_{\boldsymbol{n}} = \{\boldsymbol{u}_i\}_{i=1}^n$. The *sample center-outward distribution function* is defined as

$$\mathbf{F}^{(n)}_{\pm} := \operatorname*{argmin}_{T \in \mathcal{T}} \sum_{i=1}^{n} \Big\| \boldsymbol{z}_i - T(\boldsymbol{z}_i) \Big\|^2, \tag{3.3.1}$$

and $(n_R+1)\|\mathbf{F}^{(n)}_{\pm}(\boldsymbol{z}_i)\|$ and $\mathbf{F}^{(n)}_{\pm}(\boldsymbol{z}_i)/\|\mathbf{F}^{(n)}_{\pm}(\boldsymbol{z}_i)\|$ are called the *center-outward rank* and *center-outward sign* of $\boldsymbol{z}_i$, respectively.

Remark 3.3.1. The particular way the grid $\mathfrak{G}^d_{\boldsymbol{n}}$ is constructed here produces center-outward ranks and signs that enjoy all the properties — uniform distributions and mutual independence — that are expected from ranks and signs (see Section B.2.2 of the online Appendix). These properties, however, are not required for the finite-sample validity and asymptotic properties of the rank-based tests we are pursuing in the subsequent sections. Any sequence of grids $\mathfrak{G}^d_{\boldsymbol{n}}$, whether stochastic (defined over a different probability space than the observations) or deterministic, is fine provided that the corresponding empirical distribution converges to the spherical uniform U_d. In addition, for the reasons developed, e.g., in Hallin (2021), we deliberately only consider the spherical uniform U_d. In practice, the uniform distribution over the unit cube $[0,1]^d$ could be considered as well, yielding similar tests enjoying similar properties, with proofs following along similar lines.

The next proposition describes the Glivenko–Cantelli property of empirical center-outward distribution functions, a result we shall heavily rely on.

Proposition 3.3.1. *(Hallin, 2017, Proposition 5.1, del Barrio et al., 2018, Theorem 3.1, and Hallin et al., 2021a, Proposition 2.3) Consider the following classes of distributions:*

- *the class $\mathcal{P}_d^+$ of distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ with nonvanishing probability density, namely, with Lebesgue density f such that, for all $D > 0$ there exist constants $\lambda_{D;f} < \Lambda_{D;f} \in (0, \infty)$ such that $\lambda_{D;f} \leq f(\boldsymbol{z}) \leq \Lambda_{D;f}$ for all $\|\boldsymbol{z}\| \leq D$;*
- *the class $\mathcal{P}_d^{\#}$ of all distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ such that, denoting by $\mathbf{F}_{\pm}^{(n)}$ the sample distribution function computed from an n-tuple $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ of independent copies of $\boldsymbol{Z} \sim \mathrm{P}$,*

$$\max_{1 \leq i \leq n} \left\| \mathbf{F}_{\pm}^{(n)}(\boldsymbol{Z}_i) - \mathbf{F}_{\pm}(\boldsymbol{Z}_i) \right\| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n_R \text{ and } n_S \to \infty. \tag{3.3.2}$$

It holds that $\mathcal{P}_d^+ \subsetneq \mathcal{P}_d^{\#} \subsetneq \mathcal{P}_d^{\mathrm{ac}}$.

3.4 Rank-based dependence measures

We are now ready to present our proposed family of dependence measures based on the notions of GSCs and center-outward ranks and signs. Throughout, $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a pair of random vectors with $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$, and $(\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), (\boldsymbol{X}_{12}, \boldsymbol{X}_{22}), \ldots, (\boldsymbol{X}_{1n}, \boldsymbol{X}_{2n})$ is an n-tuple of independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$. Let $\mathbf{F}_{k,\pm}$ denote the center-outward distribution function of $\boldsymbol{X}_k$, and write $\mathbf{F}_{k,\pm}^{(n)}(\cdot)$ for the sample center-outward distribution function corresponding to $\{\boldsymbol{X}_{ki}\}_{i=1}^n$, $k = 1, 2$.

Our ideas build on Shi et al. (2021a) and, in slightly different form, also on Deb and Sen (2021), where the authors introduce a multivariate dependence measure by applying distance covariance to $\mathbf{F}_{1,\pm}(\boldsymbol{X}_1)$ and $\mathbf{F}_{2,\pm}(\boldsymbol{X}_2)$, with a sample counterpart involving $\mathbf{F}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})$ and $\mathbf{F}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})$, $i \in [\![n]\!]$. Our generalization of this particular dependence measure involves *score functions* and requires further notation. The score functions are continuous functions $J_1, J_2 : [0, 1) \to \mathbb{R}_{\geq 0}$. Classical examples include the *normal* or *van der Waerden score function* $J_{\mathrm{vdW}}(u) := \big(F_{\chi_d^2}^{-1}(u)\big)^{1/2}$ (with $F_{\chi_d^2}$ the χ_d^2 distribution function), the *Wilcoxon score* *function* $J_{\mathrm{W}}(u) := u$, and the *sign test score function* $J_{\mathrm{sign}}(u) := 1$. For $k = 1, 2$, let $\mathbf{J}_k(\boldsymbol{u}) := J_k(\|\boldsymbol{u}\|)\boldsymbol{u}/\|\boldsymbol{u}\|$ if $\boldsymbol{u} \in \mathbb{S}_{d_k} \backslash \{\mathbf{0}_{d_k}\}$ and $\mathbf{0}_{d_k}$ if $\boldsymbol{u} = \mathbf{0}_{d_k}$. Define the population and sample *scored center-outward distribution functions* as $\mathbf{G}_{k,\pm}(\cdot) := \mathbf{J}_k(\mathbf{F}_{k,\pm}(\cdot))$ and $\mathbf{G}^{(n)}_{k,\pm}(\cdot) := \mathbf{J}_k(\mathbf{F}^{(n)}_{k,\pm}(\cdot))$, respectively.

Definition 3.4.1 (Rank-based dependence measures)**.** Let J_1, J_2 be two score functions. The *(scored) rank-based version* of a dependence measure μ is obtained by applying μ to the pair $(\mathbf{G}_{1,\pm}(\boldsymbol{X}_1), \mathbf{G}_{2,\pm}(\boldsymbol{X}_2))$. For a GSC $\mu = \mu_{f_1,f_2,H}$, the rank-based version is denoted

$$\mu_{\pm}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_{\pm;J_1,J_2,f_1,f_2,H}(\boldsymbol{X}_1, \boldsymbol{X}_2) := \mu_{f_1,f_2,H}(\mathbf{G}_{1,\pm}(\boldsymbol{X}_1), \mathbf{G}_{2,\pm}(\boldsymbol{X}_2)) \tag{3.4.1}$$

and termed a *rank-based GSC* for short. The associated *rank-based SGSC* is

$$\underset{\sim}{W}^{(n)}_{\mu} = \underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} := \widehat{\mu}^{(n)}\Big(\big[\big(\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i}), \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i})\big)\big]_{i=1}^{n}; f_1, f_2, H\Big). \tag{3.4.2}$$

Remark 3.4.1. There is no immediate reason why a rank-based GSC should itself by a GSC in the sense of Definition 3.2.1. In this context, an observation of Bergsma (2006, 2011) is of interest. For distance covariance in the univariate case (equivalent to 4κ in his notation), Lemma 10 in Bergsma (2006) implies that

$$\frac{1}{16}\mu_{f_1^{\mathrm{dCov}},f_2^{\mathrm{dCov}},H_*^4}(\mathbf{G}_{X_1,\pm}(X_1), \mathbf{G}_{X_2,\pm}(X_2)) = \int (F_{(X_1,X_2)} - F_{X_1}F_{X_2})^2 \mathrm{d}F_{X_1}\mathrm{d}F_{X_2}.$$

In other words, for $d_1 = d_2 = 1$ and $J_1(u) = J_2(u) = u$, the rank-based distance covariance coincides with R of Blum et al. (1961) up to a scalar multiple. Recall that R is a GSC, but of higher order than distance covariance; see Example B.2.1(c) in Appendix B.2.1.

Plugging the center-outward ranks and signs into the multivariate dependence measures from Section 3.2 in combination with various score functions, one immediately obtains a large variety of rank-based GSCs and SGSCs, as we exemplify below. In particular, the choice $f_1 = f_1^{\mathrm{dCov}}$, $f_2 = f_2^{\mathrm{dCov}}$, $J_1(u) = J_2(u) = u$, and $H = H_*^4$ recovers the multivariate rank-based distance covariance from Shi et al. (2021a).

Example 3.4.1. Some rank-based SGSCs.

(a) Rank-based distance covariance

$$\underset{\sim}{W}^{(n)}_{\mathrm{dCov}} := \binom{n}{4}^{-1} \sum_{i_1<\cdots<i_4} h_{\mathrm{dCov}}\Big(\big(\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_1}), \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_1})\big), \ldots, \big(\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_4}), \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_4})\big)\Big)$$

with $h_{\mathrm{dCov}} := \overline{k}_{f_1^{\mathrm{dCov}}, f_2^{\mathrm{dCov}}, H_*^4}$ as given in Example 3.2.1(a). We have by definition that

$$\begin{aligned}\underset{\sim}{W}^{(n)}_{\mathrm{dCov}} = \binom{n}{4}^{-1} \sum_{i_1\neq\ldots\neq i_4} \frac{1}{4\cdot 4!}&\\
\Big[\Big\{&\big\|\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_1}) - \mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_2})\big\| - \big\|\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_1}) - \mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_3})\big\|\\
&-\big\|\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_4}) - \mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_2})\big\| + \big\|\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_4}) - \mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i_3})\big\|\Big\}\\
\times\Big\{&\big\|\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_1}) - \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_2})\big\| - \big\|\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_1}) - \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_3})\big\|\\
&-\big\|\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_4}) - \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_2})\big\| + \big\|\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_4}) - \mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i_3})\big\|\Big\}\Big];\end{aligned}$$

(b) Similarly, Hoeffding's rank-based multivariate marginal ordering D $\big($giving $\underset{\sim}{W}^{(n)}_{M}\big)$, Hoeffding's rank-based multivariate projection-averaging D $\big(\underset{\sim}{W}^{(n)}_{D}\big)$, Blum–Kiefer–Rosenblatt's rank-based multivariate projection-averaging R $\big(\underset{\sim}{W}^{(n)}_{R}\big)$, and Bergsma–Dassios–Yanagimoto's rank-based multivariate projection-averaging τ^* $\big(\underset{\sim}{W}^{(n)}_{\tau^*}\big)$ can be defined with kernels $h_M := \overline{k}_{f_1^M, f_2^M, H_*^5}$, $h_D := \overline{k}_{f_1^D, f_2^D, H_*^5}$, $h_R := \overline{k}_{f_1^R, f_2^R, H_*^6}$, and $h_{\tau^*} := \overline{k}_{f_1^{\tau^*}, f_2^{\tau^*}, H_*^4}$ as given in Example 3.2.1, respectively.

Having proposed a general class of dependence measures, we now examine, for each rank-based GSC, the five desirable properties listed in Section 3.1.2. To this end, we first introduce two regularity conditions on the score functions.

Definition 3.4.2. A score function $J : [0,1) \to \mathbb{R}_{\geq 0}$ is called *weakly regular* if it is continuous over $[0,1)$ and nondegenerate: $\int_0^1 J^2(u)\mathrm{d}u > 0$. If, moreover, J is Lipschitz-continuous, strictly monotone, and satisfies $J(0) = 0$, it is called *strongly regular*.

Proposition 3.4.1. *The normal and sign test score functions are weakly but not strongly regular; the Wilcoxon score function is strongly regular.*

Proposition 3.4.2. *Suppose the considered pair* $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *has marginal distributions* $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ *and* $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$. *Consider any rank-based GSC* $\mu_\pm := \mu_{\pm;J_1,J_2,f_1,f_2,H}$ *and its rank-based SGSC* $\underset{\sim}{W}_\mu^{(n)} := \underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)}$ *as defined in* (3.4.1) *and* (3.4.2). *Further, let* $\mu_{*\pm} := \mu_{\pm;J_1,J_2,f_1,f_2,H_*^m}$ *be an instance using the group from* (3.1.1). *Then,*

(i) (Exact distribution-freeness) Under independence of $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$*, the distribution of* $\underset{\sim}{W}_\mu^{(n)}$ *does not depend on* $\mathrm{P}_{\boldsymbol{X}_1}$ *nor* $\mathrm{P}_{\boldsymbol{X}_2}$*;*

(ii) (Transformation invariance) If the kernels f_1 *and* f_2 *are orthogonally invariant, it holds for any orthogonal matrix* $\mathbf{O}_k \in \mathbb{R}^{d_k \times d_k}$*, any vector* $\boldsymbol{v}_k \in \mathbb{R}^{d_k}$*, and any scalar* $a_k \in \mathbb{R}_{>0}$ *that* $\mu_\pm(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_\pm\big(\boldsymbol{v}_1 + a_1\mathbf{O}_1\boldsymbol{X}_1, \boldsymbol{v}_2 + a_2\mathbf{O}_2\boldsymbol{X}_2\big)$*;*

(iii) (I- and D-Consistency)

(a) $\mu_\pm$ *is I-consistent in the family*

$$\big\{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}\big|\, \mathrm{P}_{\boldsymbol{X}_k} \in \mathcal{P}_{d_k}^{\mathrm{ac}} \text{ and } \mathrm{E}\big[f_k\big([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})]_{i=1}^m\big)\big] < \infty \textit{ for } k = 1, 2\big\};$$

(b) If the pair of kernels is D-consistent in the class

$$\big\{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}\big|\, \mathrm{E}\big[f_k(\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km})\big] < \infty \textit{ for } k = 1, 2\big\}$$

(cf. Lemma 3.2.1), then $\mu_{*\pm}$ *is D-consistent in the family*

$$\mathcal{P}_{d_1,d_2,\infty}^{\mathrm{ac}} := \big\{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}\big|\, \mathrm{E}\big[f_k\big([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})]_{i=1}^m\big)\big] < \infty \textit{ for } k = 1, 2\big\} \tag{3.4.3}$$

provided that the score functions J_1 *and* J_2 *are strictly monotone;*

(iv) (Strong consistency) If $f_k\big([\mathbf{G}^{(n)}_{k,\pm}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m\big)$ *and* $f_k\big([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m\big)$ *are almost surely bounded, that is, if there exists a constant* C *(depending on* f_k*,* J_k*, and* $\mathrm{P}_{\mathbf{X}_k}$*) such that for any* n *and* $k = 1, 2$*,*

$$\mathrm{P}\Big(\,\big|f_k\big(\big[\mathbf{G}^{(n)}_{k,\pm}(\boldsymbol{X}_{ki_\ell})\big]_{\ell=1}^m\big)\big| \le C\Big) = 1 = \mathrm{P}\Big(\,\big|f_k\big(\big[\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})\big]_{\ell=1}^m\big)\big| \le C\Big),$$

and

$$(n)_m^{-1} \sum_{[i_1,\ldots,i_m]\in I_m^n} \Big| f_k\Big(\big[\mathbf{G}^{(n)}_{k,\pm}(\boldsymbol{X}_{ki_\ell})\big]_{\ell=1}^m\Big) - f_k\Big(\big[\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})\big]_{\ell=1}^m\Big)\Big| \xrightarrow{\text{a.s.}} 0, \tag{3.4.4}$$

then

$$\underset{\sim}{W}^{(n)}_{\mu} = \underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} \xrightarrow{\text{a.s.}} \mu_\pm(\boldsymbol{X}_1, \boldsymbol{X}_2). \tag{3.4.5}$$

Theorem 3.4.1 (Examples)**.** *As long as* $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}^{\#}_{d_1}$*,* $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}^{\#}_{d_2}$*, and* J_1, J_2 *are strongly regular, all the kernel functions in Example 3.2.1(a)–(e) satisfy Condition* (3.4.4)*.*

Remark 3.4.2. Unfortunately, Theorem 3.4.1 does not imply that the rank-based SGSCs with normal score functions satisfy (3.4.5) although, in view of Proposition 3.4.2(iii), their population counterparts are both I- and D-consistent within a fairly large nonparametric family of distributions. A weaker version (replacing a.s. convergence by convergence in probability) of (3.4.5) holds in the univariate case with $d_1 = d_2 = 1$ by Feuerverger (1993, Sec. 6). Consistency for normal scores, however, follows from a recent and yet unpublished result of Deb et al. (2021, Proposition 4.3), which was not available to us at the time this paper was written and which is obtained via a completely different technique.

We conclude this section with a discussion of computational issues. Two steps, in the evaluation of multivariate rank-based SGSCs, are potentially costly: (i) calculating the center-outward ranks and signs in (3.3.1), and (ii) computing a GSC $\widehat{\mu}^{(n)}(\cdot)$ with n inputs. The optimal matching problem (3.3.1) yielding $[\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i})]_{i=1}^n$ and $[\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i})]_{i=1}^n$ can be solved in $O(n^{5/2}\log(nN))$ time if the costs $\|\boldsymbol{z}_i - \boldsymbol{u}_j\|^2$, $i, j \in [\![n]\!]$ are integers bounded by N (Gabow and Tarjan, 1989); in dimension $d = 2$, this can improved to $O(n^{3/2+\delta}\log(N))$ time for some arbitrarily small constant $\delta > 0$ (Sharathkumar and Agarwal, 2012). The problem can also be solved approximately in $O(n^{3/2}\Omega(n, \epsilon, \Delta))$ time if $d \geq 3$, where

$$\Omega(n, \epsilon, \Delta) := \epsilon^{-1}\tau(n, \epsilon)\log^4(n/\epsilon)\log(\Delta)$$

depends on n, ϵ (the accuracy of the approximation) and $\Delta := \max c_{ij}/\min c_{ij}$, with $\tau(n, \epsilon)$ a small term (Agarwal and Sharathkumar, 2014). Further details are deferred to Appendix B.2.3.

Once $[\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i})]_{i=1}^n$ and $[\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i})]_{i=1}^n$ are obtained, a naïve evaluation of $\underset{\sim}{W}^{(n)}$, on the other hand, requires $O(n^m)$ operations. Great speedups are possible, however, in particular cases such as the rank-based SGSCs from Example 3.4.1. A detailed summary is provided in Proposition B.2.4 of the Appendix. The total computational complexity of the five statistics in Example 3.4.1 is given in the last three rows of Table 3.1.

3.5 Local power of rank-based tests of independence

Besides quantifying the dependence between two groups of random variables, the rank-based GSCs from Section 3.4 allow for constructing tests of the null hypothesis

$$H_0 : \boldsymbol{X}_1 \text{ and } \boldsymbol{X}_2 \text{ are mutually independent,}$$

based on a sample $(\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1n}, \boldsymbol{X}_{2n})$ of n independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$. Shi et al. (2021a), and, in a slightly different manner, Deb and Sen (2021), studied the particular case of a test based on the Wilcoxon version of the rank-based distance covariance $\underset{\sim}{W}^{(n)}_{\text{dCov}}$. Among other results, they derive the limiting null distribution of $\underset{\sim}{W}^{(n)}_{\text{dCov}}$, using combinatorial limit theorems and "brute-force" calculation of permutation statistics. Although this led to a fairly general combinatorial non-central limit theorem (Shi et al., 2021a, Theorems 4.1 and 4.2), the derivation is not intuitive and difficult to generalize. In contrast, in this paper, we take a new and more powerful approach to the asymptotic analysis of rank-based SGSCs, which resolves the following three main issues:

(i) Intuitively, the asymptotic behavior of rank-based dependence measures follows from that of their Hájek *asymptotic representations*, which are oracle versions in which the observations are transformed using the unknown actual center-outward distribution function $\mathbf{F}_{\pm}$ rather than its sample version $\mathbf{F}_{\pm}^{(n)}$. Here, we show the correctness of this intuition by proving asymptotic equivalence between rank-based SGSCs and their oracle versions.

(ii) Previous work does not perform any power analysis for the new rank-based tests. Here, we fill this gap by proving that these tests have nontrivial power in the context of the class of quadratic mean differentiable alternatives (Lehmann and Romano, 2005, Def. 12.2.1).

(iii) Finally, our rank-based tests allow for the incorporation of score functions, which may improve their performance.

This novel approach rests on a generalization of the classical Hájek representation method (Hájek and Šidák, 1967) to the multivariate setting of center-outward ranks and signs, which simplifies the derivation of asymptotic null distributions and, via a nontrivial use of Le Cam's third lemma for non-normal limits, enables our local power analysis.

3.5.1 Asymptotic representation

In order to develop our multivariate asymptotic representation, we first introduce formally the oracle counterpart to the rank-based SGSC $\underset{\sim}{W}_{\mu}^{(n)}$.

Definition 3.5.1 (Oracle rank-based SGSCs)**.** The oracle version of the rank-based SGSC $\underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}}$ associated with the GSC $\mu = \mu_{f_1,f_2,H}$ is

$$W^{(n)}_{\mu} = W^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} := \widehat{\mu}^{(n)}\Big(\big[\big(\mathbf{G}_{1,\pm}(\boldsymbol{X}_{1i}), \mathbf{G}_{2,\pm}(\boldsymbol{X}_{2i})\big)\big]_{i=1}^{n}; f_1, f_2, H\Big).$$

Note that the oracle $W^{(n)}_{\mu}$ cannot be computed from the observations as it involves the population scored center-outward distribution functions $\mathbf{G}_{1,\pm}$ and $\mathbf{G}_{2,\pm}$. However, the limiting null distribution of $W^{(n)}$, unlike that of $\underset{\sim}{W}^{(n)}$, follows from standard theory for degenerate U-statistics (Serfling, 1980, Chap. 5.5.2). This point can be summarized as follows.

Proposition 3.5.1. *Let $\mu = \mu_{f_1,f_2,H^m_*}$ be a GSC with $m \geq 4$. Let the kernels f_1, f_2 and the score functions J_1, J_2 satisfy*

$$0 < \mathrm{Var}(g_k(\boldsymbol{W}_{k1}, \boldsymbol{W}_{k2})) < \infty, \quad k = 1, 2, \tag{3.5.1}$$

where $\boldsymbol{W}_{ki} := \mathbf{J}_k(\boldsymbol{U}_{ki})$ with $(\boldsymbol{U}_{1i}, \boldsymbol{U}_{2i})$, $i \in [\![m]\!]$ independent and distributed according to the product of spherical uniform distributions $\mathrm{U}_{d_1} \otimes \mathrm{U}_{d_2}$,

$$g_k(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}) := \mathrm{E}\Big[2 f_{k,H^m_*}\Big(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}, \boldsymbol{W}_{k3}, \boldsymbol{W}_{k4}, \ldots, \boldsymbol{W}_{km}\Big)\Big], \tag{3.5.2}$$

and $f_{k,H^m_*} := \sum_{\sigma \in H^m_*} \mathrm{sgn}(\sigma) f_k(\boldsymbol{x}_{k\sigma(1)}, \ldots, \boldsymbol{x}_{k\sigma(m)})$, $k = 1, 2$. Then, under the null hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}^{\mathrm{ac}}_{d_1}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}^{\mathrm{ac}}_{d_2}$ are independent,

$$nW^{(n)}_{\mu} = nW^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H^m_*}} \rightsquigarrow \sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1),$$

where $[\xi_v]_{v=1}^{\infty}$ are independent standard Gaussian random variables and $[\lambda_{\mu,v}]_{v=1}^{\infty}$ are the nonzero eigenvalues of the integral equation

$$\mathrm{E}\big[g_1(\boldsymbol{w}_{11}, \boldsymbol{W}_{12}) g_2(\boldsymbol{w}_{21}, \boldsymbol{W}_{22}) \psi(\boldsymbol{W}_{12}, \boldsymbol{W}_{22})\big] = \lambda \psi(\boldsymbol{w}_{11}, \boldsymbol{w}_{21}). \tag{3.5.3}$$

The tests we are considering reject for large values of test statistics that estimate a nonnegative (I- and D-)consistent dependence measure. In all these tests

$$\text{all eigenvalues of the integral equation (3.5.3) are non-negative.} \tag{3.5.4}$$

However, it should be noted that, in view of the following multivariate representation result, a valid test of H_0 can be implemented also when (3.5.4) does not hold.

Theorem 3.5.1 (Multivariate Hájek representation)**.** *Let f_1, f_2 be kernel functions of order $m \geq 4$, and let J_1, J_2 be weakly regular score functions. Writing $\mathrm{U}_{d_k}^{(n)}$ for the discrete uniform distribution over the grid $\mathfrak{G}_{\boldsymbol{n}}^{d_k}$, let $\boldsymbol{W}_{ki}^{(n)} := \mathbf{J}_k(\boldsymbol{U}_{ki}^{(n)})$ where $(\boldsymbol{U}_{1i}^{(n)}, \boldsymbol{U}_{2i}^{(n)})$ for $i \in [\![m]\!]$ are independent with distribution $\mathrm{U}_{d_1}^{(n)} \otimes \mathrm{U}_{d_2}^{(n)}$. Define g_k, $k = 1, 2$, as in (3.5.2), and*

$$g_k^{(n)}(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}) := \mathrm{E}\Big[2 f_{k,H_*^m}\Big(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}, \boldsymbol{W}_{k3}^{(n)}, \boldsymbol{W}_{k4}^{(n)}, \ldots, \boldsymbol{W}_{km}^{(n)}\Big)\Big], \quad k = 1, 2. \tag{3.5.5}$$

Assume that

$$f_k \text{ and } g_k \text{ are Lipschitz-continuous,} \quad g_k^{(n)} \text{ converges uniformly to } g_k, \tag{3.5.6}$$
$$\sup_{i_1,\ldots,i_m \in [\![m]\!]} \mathrm{E}[f_k([\boldsymbol{W}_{ki_\ell}]_{\ell=1}^m)^2] < \infty, \quad \text{and} \quad \int_0^1 J_k^2(u)\mathrm{d}u < \infty, \quad k = 1, 2.$$

Then, under the hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ are independent, the rank-based SGSC $\underset{\sim}{W}_{\mu}^{(n)} = \underset{\sim}{W}_{J_1,J_2,\mu}^{(n)}$ associated to the GSC $\mu = \mu_{f_1,f_2,H_*^m}$ is asymptotically equivalent to its oracle version $W_\mu^{(n)}$, i.e., $\underset{\sim}{W}_{\mu}^{(n)} - W_\mu^{(n)} = o_{\mathrm{P}}(n^{-1})$ as $n_R, n_S \to \infty$.

Theorem 3.5.2. *The conclusion of Theorem 3.5.1 still holds with (3.5.6) replaced by*

$$f_k \text{ is uniformly bounded, and almost everywhere continuous,} \quad k = 1, 2. \tag{3.5.7}$$

Proposition 3.5.2 (Examples)**.** *If $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ is independent of $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ and J_1, J_2 are weakly regular, then the kernel functions from Example 3.2.1(b)–(e) satisfy (3.5.1), (3.5.4), and (3.5.7). If, moreover, J_1, J_2 are square-integrable (viz., $\int_0^1 J_k^2(u)\mathrm{d}u < \infty$ for $k = 1, 2$), then (3.5.1), (3.5.4), and (3.5.6) hold also for the kernels in Exam-*

ple 3.2.1(a).

Corollary 3.5.1 (Limiting null distribution)**.** *Suppose the conditions in Proposition 3.5.1 and Theorem 3.5.1 hold. Then, for $\mu = \mu_{f_1,f_2,H_*^m}$ with $m \geq 4$, under the hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ are independent,*

$$n\underset{\sim}{W}_{\mu}^{(n)} = n\underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H_*^m}}^{(n)} \rightsquigarrow \sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1) \tag{3.5.8}$$

with $[\lambda_{\mu,v}]_{v=1}^{\infty}$ and $[\xi_v]_{v=1}^{\infty}$ as defined in Proposition 3.5.1.

Remark 3.5.1. Corollary 3.5.1 gives no rate, i.e., no Berry–Esséen type bound for the convergence in (3.5.8). Indeed, deriving such bounds in the present context is quite challenging. Results for the univariate case with $d_1 = d_2 = 1$ were established for simpler statistics such as Spearman's ρ and Kendall's τ by Koroljuk and Borovskich (1994, Chap. 6.2) and, more recently, by Pinelis and Molzon (2016). Extending these results to the multivariate measure-transportation-based ranks considered here is highly nontrivial and requires properties of empirical transports that have not yet been obtained. This pertains, in particular, to working out the rate of convergence in the Glivenko–Cantelli result for the center-outward distribution function given in (3.3.2); an open problem in the recent survey by Hallin (2021, Section 5).

For any significance level $\alpha \in (0, 1)$, define the quantile

$$q_{\mu,1-\alpha} := \inf\Big\{x \in \mathbb{R} : \mathrm{P}\Big(\sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1) \leq x\Big) \geq 1 - \alpha\Big\}, \tag{3.5.9}$$

where $[\lambda_{\mu,v}]_{v=1}^{\infty}$ and $[\xi_v]_{v=1}^{\infty}$ are as in Proposition 3.5.1. Let $\underset{\sim}{W}_{\mu}^{(n)}$ be as in Theorem 3.5.1, and define the test

$$\mathsf{T}_{\mu,\alpha}^{(n)} := \mathbb{1}\big(n\underset{\sim}{W}_{\mu}^{(n)} > q_{\mu,1-\alpha}\big).$$

The next proposition summarizes the asymptotic validity and properties of this test.

Proposition 3.5.3 (Uniform validity and consistency)**.** *Let J_1, J_2 be weakly regular score functions, and let $\mu = \mu_{f_1,f_2,H_*^m}$ be a GSC with $m \geq 4$ such that Conditions* (3.5.1) *and one of* (3.5.6) *and* (3.5.7) *hold. Then,*

(i) $\lim_{n\to\infty} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = \alpha$ *for any* $\mathrm{P} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$*, i.e., for* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *independent with* $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ *and* $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$*;*

(ii) it follows from Proposition 3.4.2(i) that $\lim_{n\to\infty} \sup_{\mathrm{P}\in\mathcal{P}_{d_1}^{\#}\otimes\mathcal{P}_{d_2}^{\#}} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = \alpha$*;*

(iii) if, moreover, the pair of kernels (f_1, f_2) *is D-consistent,* J_1, J_2 *are strictly monotone, and* (3.4.5) *holds,* $\lim_{n\to\infty} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = 1$ *for any fixed alternative* $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2,\infty}^{\mathrm{ac}}$ *as defined in* (3.4.3)*.*

3.5.2 Local power analysis

In this section, we conduct local power analyses of the proposed tests for quadratic mean differentiable classes of alternatives (Lehmann and Romano, 2005, Def. 12.2.1), for which we establish nontrivial power in $n^{-1/2}$ neighborhoods. We begin with a model $\{q_{\boldsymbol{X}}(\boldsymbol{x};\delta)\}_{|\delta|<\delta^*}$ with $\delta^* > 0$, under which $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ has Lebesgue-density $q_{\boldsymbol{X}}(\boldsymbol{x};\delta) = q_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}\big((\boldsymbol{x}_1,\boldsymbol{x}_2);\delta\big)$, with $q_{\boldsymbol{X}_1}(\boldsymbol{x}_1;\delta)$ and $q_{\boldsymbol{X}_2}(\boldsymbol{x}_2;\delta)$ being the marginal densities. We then make the following assumptions.

Assumption 3.5.1.

(i) Dependence of $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$*:* $q_{\boldsymbol{X}}(\boldsymbol{x};\delta) = q_{\boldsymbol{X}_1}(\boldsymbol{x}_1;\delta)q_{\boldsymbol{X}_2}(\boldsymbol{x}_2;\delta)$ *holds if and only if* $\delta = 0$*.*

(ii) The family $\{q_\delta(\boldsymbol{x})\}_{|\delta|<\delta^*}$ *is quadratic mean differentiable at* $\delta = 0$ *with score function* $\dot{\ell}(\cdot;0)$*, that is,*

$$\int \left(\sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};\delta)} - \sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};0)} - \frac{1}{2}\delta\dot{\ell}(\boldsymbol{x};0)\sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};0)}\right)^2 \mathrm{d}\boldsymbol{x} = o(\delta^2) \quad \textit{as} \;\; \delta \to 0.$$

(iii) The Fisher information is positive, i.e., $\mathcal{I}_{\boldsymbol{X}}(0) := \int \{\dot{\ell}(\boldsymbol{x};0)\}^2 q_{\mathbf{X}}(\boldsymbol{x},0)\mathrm{d}\boldsymbol{x} > 0$*; of note, Assumption 3.5.1(ii) implies that* $\mathcal{I}_{\boldsymbol{X}}(0) < \infty$ *and* $\int \dot{\ell}(\boldsymbol{x};0)q_{\mathbf{X}}(\boldsymbol{x},0)\mathrm{d}\boldsymbol{x} = 0$*.*

(iv) The score function $\dot{\ell}(\boldsymbol{x};0)$ *is not additively separable, i.e., there do not exist functions* h_1 *and* h_2 *such that* $\dot{\ell}(\boldsymbol{x};0) = h_1(\boldsymbol{x}_1) + h_2(\boldsymbol{x}_2)$.

Remark 3.5.2. For the sake of simplicity, we have restricted ourselves to one-parameter classes. Analogous results hold for families indexed by a multivariate parameter $\boldsymbol{\delta}$.

For a local power analysis, we consider a sequence of local alternatives obtained as

$$H_1^{(n)}(\delta_0) : \delta = \delta^{(n)}, \quad \text{where } \delta^{(n)} := n^{-1/2}\delta_0 \tag{3.5.10}$$

with some constant $\delta_0 \neq 0$. In this local model, testing the null hypothesis of independence reduces to testing $H_0 : \delta_0 = 0$ versus $H_1 : \delta_0 \neq 0$.

Theorem 3.5.3 (Power analysis)**.** *Consider a GSC* $\mu = \mu_{f_1,f_2,H_*^m}$ *with* $m \geq 4$ *and kernel functions* f_1, f_2 *picked from Example 3.2.1. Assume that* J_1, J_2 *are weakly regular score functions that satisfy the assumptions of Proposition 3.5.2. Then if Assumption 3.5.1 holds, for any* $\beta > 0$*, there exists a constant* $C_\beta > 0$ *depending only on* β *such that, as long as* $|\delta_0| > C_\beta$, $\lim_{n\to\infty} \mathrm{P}\big\{\mathsf{T}_{\mu,\alpha}^{(n)} = 1 \big| H_1^{(n)}(\delta_0)\big\} \geq 1 - \beta$.

Following the arguments from the proof of Theorem 3.5.3, one should be able to obtain similar local power results for the original (non-rank-based) tests associated with the kernels listed in Example 3.2.1. However, to the best of our knowledge, this analysis has not been performed in the literature, except for $d_1 = d_2 = 1$ where results can be found, e.g., in Dhar et al. (2016) and Shi et al. (2021b). We also emphasize that, although Theorem 3.5.3 only considers the specific cases listed also in Example 3.4.1, the proof technique applies more generally. We refrain, however, from stating a more general version of Theorem 3.5.3 as this would require a number of tedious technical conditions.

Combined with the following result, Theorem 3.5.3 yields nontrivial power of the proposed tests in $n^{-1/2}$ neighborhoods of $\delta = 0$.

Table 3.1: Properties of the center-outward GSCs in Example 3.4.1 with weakly regular score functions J_k.

		$\underset{\sim}{W}_{\mu}^{(n)}$	$\underset{\sim}{W}_{\mathrm{dCov}}^{(n)}$	$\underset{\sim}{W}_{M}^{(n)}$	$\underset{\sim}{W}_{D}^{(n)}$	$\underset{\sim}{W}_{R}^{(n)}$	$\underset{\sim}{W}_{\tau^*}^{(n)}$
(1)	Distribution-freeness		$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$ [(a)]	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$
(2)	Transformation invariance		Orthogonal transf., shifts, and global scales	Shifts and global scales	Orthogonal transf., shifts, and global scales	Orthogonal transf., shifts, and global scales	Orthogonal transf., shifts, and global scales
(3)	D-consistency		J_k strictly monotone and integrable, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$ [(b)]	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$
(3')	Consistency of test		J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2}^{\#}$ [(c)]	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2}^{\#}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2}^{\#}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2}^{\#}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2}^{\#}$
(4)	Efficiency		J_k square-integrable	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)
(5)	Exact	$d_1 \vee d_2 = 2$	$O(n^2)$	$O(n^{3/2+\delta} \log N)$ [(d)]	$O(n^3)$	$O(n^4)$	$O(n^4)$
		$d_1 \vee d_2 = 3$	$O(n^{5/2} \log(nN))$ [(d)]	$O(n^{5/2} \log(nN))$	$O(n^3)$	$O(n^4)$	$O(n^4)$
	Fast approximation		$O(n^{3/2}\Omega \vee nK \log n)$ [(d)]	$O(n^{3/2}\Omega)$	$O(n^{3/2}\Omega \vee nK \log n)$	$O(n^{3/2}\Omega \vee nK \log n)$	$O(n^{3/2}\Omega \vee nK \log n)$

(a) $\mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$ is the family of all $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}$ such that $\boldsymbol{X}_1, \boldsymbol{X}_2$ independent, $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$

(b) $\mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$ is the family of all absolutely continuous distributions on $\mathbb{R}^{d_1+d_2}$

(c) $\mathcal{P}_{d_1,d_2}^{\#} := \{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}} | \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\#}, \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\#}\}$

(d) Here we assume without loss of generality that $c_{ij},\ i,j \in [\![n]\!]$ are all integers and bounded by integer N, δ is some arbitrarily small constant, Ω is defined as $\epsilon^{-1}\tau(n,\epsilon)\log^4(n/\epsilon)\log(\max c_{ij}/\min c_{ij})$, and K is sufficiently large; as usual, $q_1 \vee q_2$ stands for the minimum of two quantities q_1 and q_2. Also refer to Propositions B.2.3 and B.2.4 in Section B.2 of the appendix.

Theorem 3.5.4. *Let Assumption 3.5.1 hold. Then, for any $\beta > 0$ such that $\alpha + \beta < 1$, there exists an absolute constant $c_\beta > 0$ such that, as long as $|\delta_0| \le c_\beta$,*

$$\inf_{\overline{\mathsf{T}}_\alpha^{(n)} \in \mathcal{T}_\alpha^{(n)}} \mathrm{P}\big\{\overline{\mathsf{T}}_\alpha^{(n)} = 0 \big| H_1^{(n)}(\delta_0)\big\} \ge 1 - \alpha - \beta$$

for all sufficiently large n. Here the infimum is taken over the class $\mathcal{T}_\alpha^{(n)}$ of all size-α tests.

Table 3.1 summarizes our results for the rank-based SGSCs from Example 3.4.1 by giving an overview of the five properties listed in the Introduction. It also indicates consistency of the tests. In all cases, it is assumed that the score functions involved are weakly regular.

3.5.3 Examples in the quadratic mean differentiable class

This section presents two specific examples in the quadratic mean differentiable class that satisfy Assumption 3.5.1. First, we consider parametrized families that extend the bivariate *Konijn alternatives* (Konijn, 1956). These alternatives are classical in the context of testing for multivariate independence and have also been considered by Gieser (1993), Gieser and Randles (1997), Taskinen et al. (2003, 2004), Taskinen et al. (2005), and Hallin and Paindaveine (2008).

Konijn families are constructed as follows. Let $\boldsymbol{X}_1^* \sim \mathrm{P}_{\boldsymbol{X}_1^*} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2^* \sim \mathrm{P}_{\boldsymbol{X}_2^*} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ be two (without loss of generality) mean zero (unobserved) independent random vectors with densities q_1 and q_2, respectively. Let $\mathbf{G}_{1,\pm}^*$ and $\mathbf{G}_{2,\pm}^*$ denote their respective population scored center-outward distribution functions, $\mathrm{P}_{\boldsymbol{X}^*} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}}$ their joint distribution, $q_{\boldsymbol{X}^*}(\boldsymbol{x}) = q_{\boldsymbol{X}^*}((\boldsymbol{x}_1, \boldsymbol{x}_2)) = q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)$ their joint density. Define, for $\delta \in \mathbb{R}$,

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} := \begin{pmatrix} \mathbf{I}_{d_1} & \delta\,\mathbf{M}_1 \\ \delta\,\mathbf{M}_2 & \mathbf{I}_{d_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1^* \\ \boldsymbol{X}_2^* \end{pmatrix} = \mathbf{A}_\delta \begin{pmatrix} \boldsymbol{X}_1^* \\ \boldsymbol{X}_2^* \end{pmatrix} = \mathbf{A}_\delta \boldsymbol{X}^*, \tag{3.5.11}$$

where $\mathbf{M}_1 \in \mathbb{R}^{d_1 \times d_2}$ and $\mathbf{M}_2 \in \mathbb{R}^{d_2 \times d_1}$ are two deterministic matrices. For $\delta = 0$, the matrix $\mathbf{A}_\delta$ is the identity and, thus, invertible. By continuity, $\mathbf{A}_\delta$ is also invertible for δ

in a sufficiently small neighborhood Θ of 0. For $\delta \in \Theta$, the density of $\boldsymbol{X}$ can be expressed as $q_{\boldsymbol{X}}(\boldsymbol{x};\delta) = \left|\det(\mathbf{A}_\delta)\right|^{-1} q_{\boldsymbol{X}^*}(\mathbf{A}_\delta^{-1}\boldsymbol{x})$, which is differentiable with respect to δ. The following additional assumptions will be made on the generating scheme (3.5.11).

Assumption 3.5.2.

(i) *The distributions of $\boldsymbol{X}$ have a common support for all $\delta \in \Theta$. Without loss of generality, we assume $\boldsymbol{\mathcal{X}} := \{\boldsymbol{x} : q_{\boldsymbol{X}}(\boldsymbol{x};\delta) > 0\}$ does not depend on δ.*

(ii) *The map $\boldsymbol{x} \mapsto \sqrt{q_{\boldsymbol{X}^*}(\boldsymbol{x})}$ is continuously differentiable.*

(iii) *The Fisher information $\mathcal{I}_{\boldsymbol{X}}(0) := \int \{\dot{\ell}(\boldsymbol{x};0)\}^2 q_{\boldsymbol{X}}(\boldsymbol{x};0)\mathrm{d}\boldsymbol{x}$ of $\boldsymbol{X}$ relative to δ at $\delta = 0$ is strictly positive and finite.*

Example 3.5.1.

(i) Suppose $\boldsymbol{X}_1^*$ and $\boldsymbol{X}_2^*$ are elliptical with centers $\mathbf{0}_{d_1}$ and $\mathbf{0}_{d_2}$ and covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively, that is, $q_k(\boldsymbol{x}_k) \propto \phi_k\Big(\boldsymbol{x}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{x}_k\Big)$, $k = 1, 2$, where ϕ_k is such that $\mathrm{Var}(\boldsymbol{X}_k^*) = \boldsymbol{\Sigma}_k$ and $\mathrm{E}\left[\|\boldsymbol{Z}_k^*\|^2 \rho_k(\|\boldsymbol{Z}_k^*\|^2)^2\right] < \infty$, $k = 1, 2$ where $\boldsymbol{Z}_k^*$ has density function proportional to $\phi_k(\|\boldsymbol{z}_k\|^2)$ and $\rho_k(t) := \phi_k'(t)/\phi_k(t)$. Then Assumption 3.5.2 is satisfied for any $\mathbf{M}_1, \mathbf{M}_2$ such that $\boldsymbol{\Sigma}_1 \mathbf{M}_2^\top + \mathbf{M}_1 \boldsymbol{\Sigma}_2 \neq \mathbf{0}$.

(ii) As a specific example of (i), if $\boldsymbol{X}_1^*$ and $\boldsymbol{X}_2^*$ are centered multivariate normal or follow centered multivariate t-distributions with degrees of freedom strictly greater than two, then Assumption 3.5.2 is satisfied for any $\mathbf{M}_1, \mathbf{M}_2$ such that $\boldsymbol{\Sigma}_1 \mathbf{M}_2^\top + \mathbf{M}_1 \boldsymbol{\Sigma}_2 \neq \mathbf{0}$.

Next, consider the following mixture model extending the alternatives treated in Dhar et al. (2016, Sec. 3). Let q_1 and q_2 be fixed (Lebesgue-)density functions for $\boldsymbol{X}_1$ an $\boldsymbol{X}_2$, respectively. The joint density of $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ under independence is $q_1 q_2$. Letting $q^* \neq q_1 q_2$ denote a fixed joint density, mixture alternatives indexed by $\delta \in [0, 1]$ are defined as $q_{\boldsymbol{X}}(\boldsymbol{x};\delta) := (1-\delta) q_1 q_2 + \delta q^*$.

Assumption 3.5.3. *It is assumed that*

(i) $(1+\delta^*)q_1q_2 - \delta^* q^*$ *is a bonafide joint density for some* $\delta^* > 0$*;*

(ii) q^* *and* q_1q_2 *are mutually absolutely continuous;*

(iii) the function $\delta \mapsto \sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};\delta)}$ *is continuously differentiable in some neighborhood of* 0*;*

(iv) the Fisher information $\mathcal{I}_{\boldsymbol{X}}(\delta) := \int (q^* - q_1q_2)^2/\{(1-\delta)q_1q_2 + \delta q^*\}\mathrm{d}\boldsymbol{x}$ *of* $\boldsymbol{X}$ *relative to* δ *is finite, strictly positive, and continuous at* $\delta = 0$*;*

(v) $\dot{\ell}(\boldsymbol{x};0) = q^*(\boldsymbol{x})/\{q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)\} - 1$ *is not additively separable.*

Example 3.5.2. If $q_k(\boldsymbol{x}_k) = 1$ for $\boldsymbol{x}_k \in [0,1]^{d_k}$, $k = 1,2$, and $q^*(\boldsymbol{x}) \not\equiv 1$ is continuous and supported on $[0,1]^{d_1+d_2}$, then Assumption 3.5.3 holds.

Proposition 3.5.4. *Assumption 3.5.1 is satisfied by the Konijn alternatives under Assumption 3.5.2, and by the mixture alternatives under Assumption 3.5.3.*

3.5.4 Numerical experiments

Extensive simulations of Shi et al. (2021a) give evidence for the superiority, under non-Gaussian densities, of the Wilcoxon versions of our tests over the original distance covariance tests. That superiority is more substantial when non-Wilcoxon scores, such as the Gaussian ones, are considered (Figure 3.4). In view of these results, there is little point in pursuing simulations with non-Gaussian densities, and we instead focus on Gaussian cases (Figures 3.1–3.3) to study the impact on finite-sample performance of the dimensions d_1 and d_2, sample size n, and within- and between-sample correlations.

Example 3.5.3. The data are a sample of n independent copies of the multivariate normal vector $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ in $\mathbb{R}^{d_1+d_2}$, with mean zero and covariance matrix $\boldsymbol{\Sigma}$, where

$$\Sigma_{ij} = \Sigma_{ji} = \begin{cases} 1, & i = j, \\ \tau, & i = 1,\, j = 2, \\ \rho, & i = 1,\, j = d_1 + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Here τ characterizes the within-group correlation and we consider (a) $\tau = 0$, (b) $\tau = 0.5$, and (c) $\tau = 0.9$. Independence holds if and only if ρ, a between-group correlation, is zero.

Example 3.5.4. The data are n independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ with $X_{1i} = Q_{t(1)}(\Phi(X^*_{1i}))$ and $X_{2j} = Q_{t(1)}(\Phi(X^*_{2j}))$ for $i \in [\![d_1]\!]$ and $j \in [\![d_2]\!]$; here $Q_{t(1)}$ denotes the quantile function of the standard Cauchy distribution and $(\boldsymbol{X}^*_1, \boldsymbol{X}^*_2)$ is generated according to Example 3.5.3(b).

We compare the empirical performance of the following five tests:

(i) permutation test using the original distance covariance (Székely and Rizzo, 2013);

(ii) permutation test applying original distance covariance to marginal ranks (Lin, 2017);

(iii) center-outward rank-based distance covariance test with Wilcoxon scores and critical values from the asymptotic distribution (Shi et al., 2021a);

(iv) new center-outward rank-based distance covariance test with normal scores and critical values from the asymptotic distribution;

(v) likelihood ratio test in the Gaussian model (Anderson, 2003, Chap. 9.3.3 & 8.4.4).

The parametric test (v) is tailored for Gaussian densities and plays the role of a benchmark. Unsurprisingly, in the Gaussian experiments in Figures 3.1–3.3, it uniformly outperforms tests (i)-(iv). See Figure 3.4 for its unsatisfactory performance for non-Gaussian densities.

Figures 3.1–3.4 report empirical powers (rejection frequencies) of these five tests, based on $1,000$ simulations with nominal significance level 0.05, dimensions $d_1 = d_2 \in \{2, 3, 5, 7\}$, and sample size $n \in \{216, 432, 864, 1728\}$. The parameter ρ in the true covariance matrix takes values $\rho \in \{0, 0.005, \ldots, 0.15\}$. The critical values for tests (i) and (ii) were computed on the basis of n random permutations. For tests (iii) and (iv), to determine the critical values from the asymptotic distribution given in Corollary 3.5.1, we numerically compute the eigenvalues by adopting the same strategy as in Shi et al. (2021a, Sec. 5.2); see also Lyons (2013, p. 3291).

It is evident from Figure 3.4 that, in non-Gaussian experiments, the potential benefits of rank-based tests are huge, particularly so when Gaussian scores are adopted (note the very severe bias of the Gaussian likelihood ratio test as d increases). In Gaussian experiments, the performance of the normal score–based test (iv) is uniformly better than that of its Wilcoxon score counterpart (iii); that superiority increases with the dimension and decreases with the within-group dependence τ. The superiority of both center-outward rank-based tests (iii) and (iv) over the traditional distance covariance one and its marginal rank version is quite significant for high values of the within-group correlation τ.

The way the normal-score rank-based test (and also the Wilcoxon-score one) outperforms the original distance covariance test may come as a surprise. However, the original distance covariance does not yield a Gaussian parametric test but rather a nonparametric test for which there is no reason to expect superiority over its rank-based versions in Gaussian settings. In a different context, we have long been used to the celebrated Chernoff–Savage phenomenon that normal-score rank statistics may (uniformly) outperform their pseudo-Gaussian counterparts (Chernoff and Savage, 1958). This is best known in the context of two-sample location problems; see, however, Hallin (1994), Hallin and Paindaveine (2008), and Deb et al. (2021) for Chernoff–Savage results for linear time series (traditional univariate ranks and correlogram-based pseudo-Gaussian procedures) and vector independence (Mahalanobis ranks and signs under elliptical symmetry and Wilks' test as the pseudo-Gaussian procedure; measure-transportation-based ranks under elliptical symmetry or independent component assumptions). Although the present context is different, their superiority is another example in which restricting to rank-based methods brings distribution-freeness at no substantial cost in terms of efficiency/power.

3.6 Conclusion

This paper provides a general framework for specifying dependence measures that leverage the new concept of center-outward ranks and signs. The associated independence tests have the strong appeal of being fully distribution-free. Via the use of a flexible class of generalized symmetric covariances and the incorporation of score functions, our framework allows one to construct a variety of consistent dependence measures. This, as our numerical experiments demonstrate, can lead to significant gains in power.

The theory we develop facilitates the derivation of asymptotic distributions yielding easily computable approximate critical values. The key result is an asymptotic representation that also allows us to establish, for the first time, a nontrivial local power result for tests of vector independence based on center-outward ranks and signs.

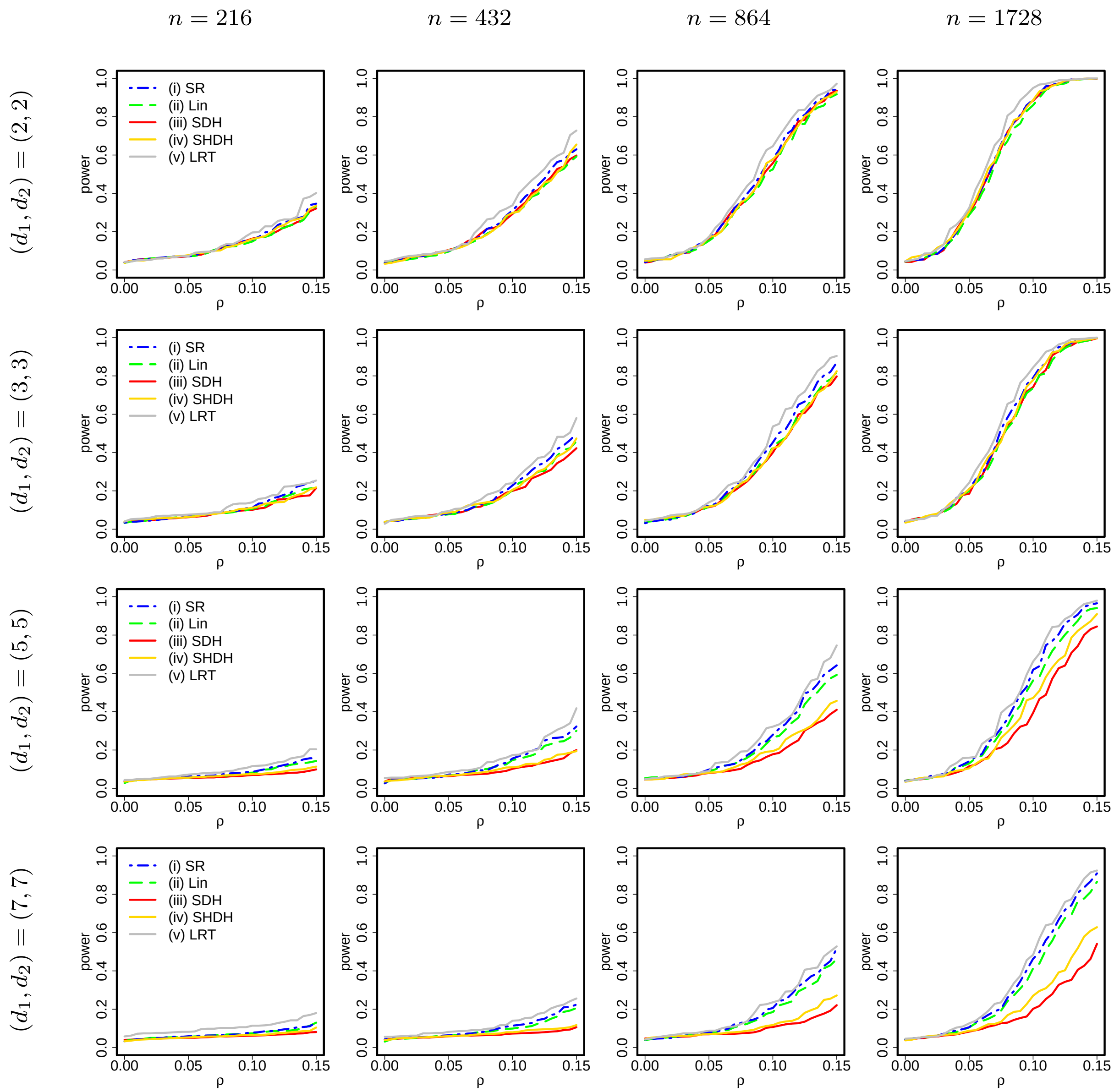


Figure 3.1: Empirical powers of the five competing tests in Example 3.5.3(a) ($\tau = 0$, no within-group correlation). The y-axis represents rejection frequencies based on 1,000 replicates, the x-axis represents ρ (the between-group correlation), and the blue, green, red, and gold lines represent the performance of (i) Szekely and Rizzo's original distance covariance test, (ii) Lin's marginal rank version of the distance covariance test, (iii) Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test, (iv) the center-outward normal-score version of the distance covariance test, and (v) the likelihood ratio test, respectively.

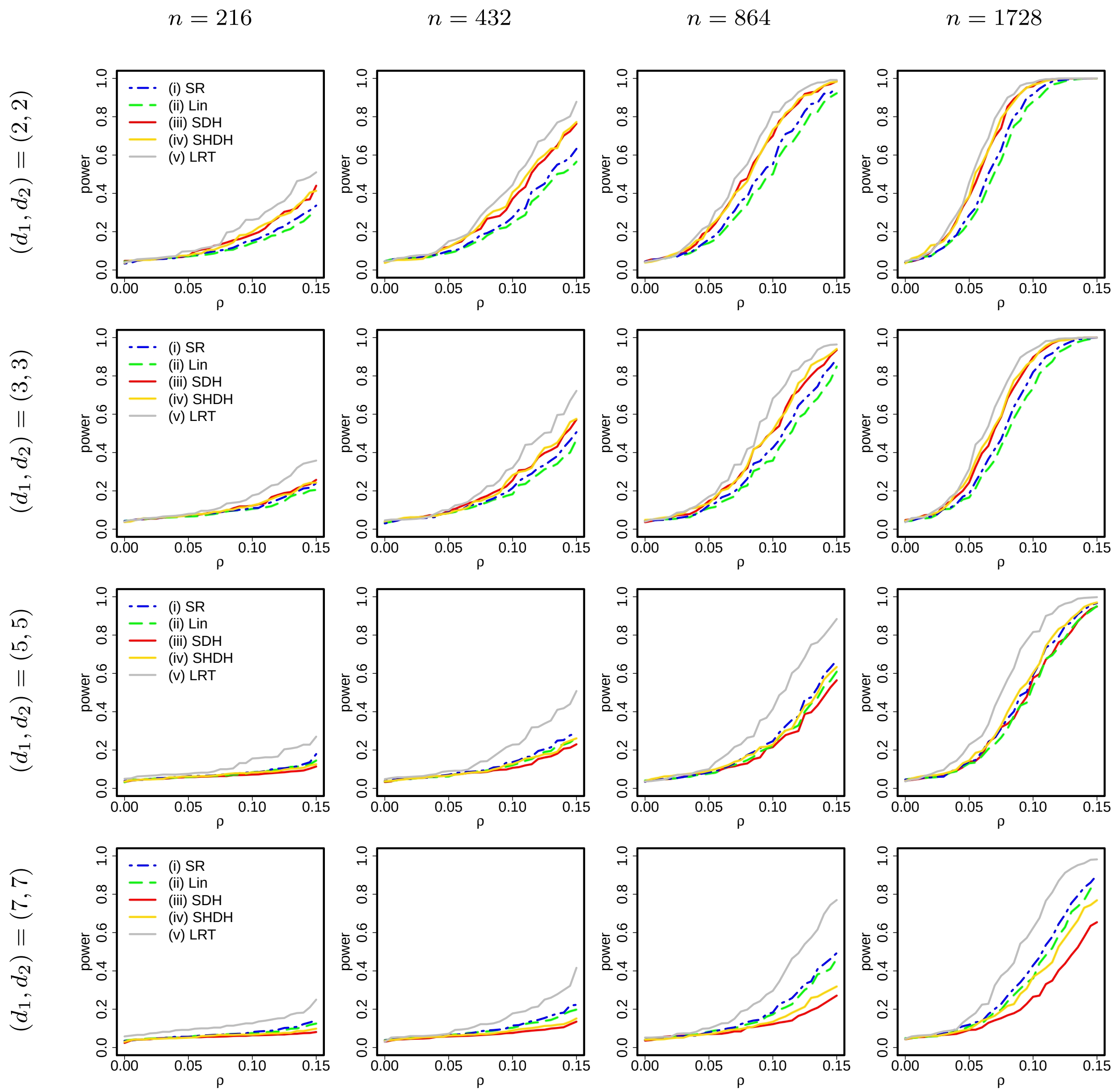


Figure 3.2: Empirical powers of the five competing tests in Example 3.5.3(b) ($\tau = 0.5$, moderate within-group correlation). The y-axis represents rejection frequencies based on 1,000 replicates, the x-axis represents ρ (the between-group correlation), and the blue, green, red, and gold lines represent the performance of (i) Szekely and Rizzo's original distance covariance test, (ii) Lin's marginal rank version of the distance covariance test, (iii) Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test, (iv) the center-outward normal-score version of the distance covariance test, and (v) the likelihood ratio test, respectively.

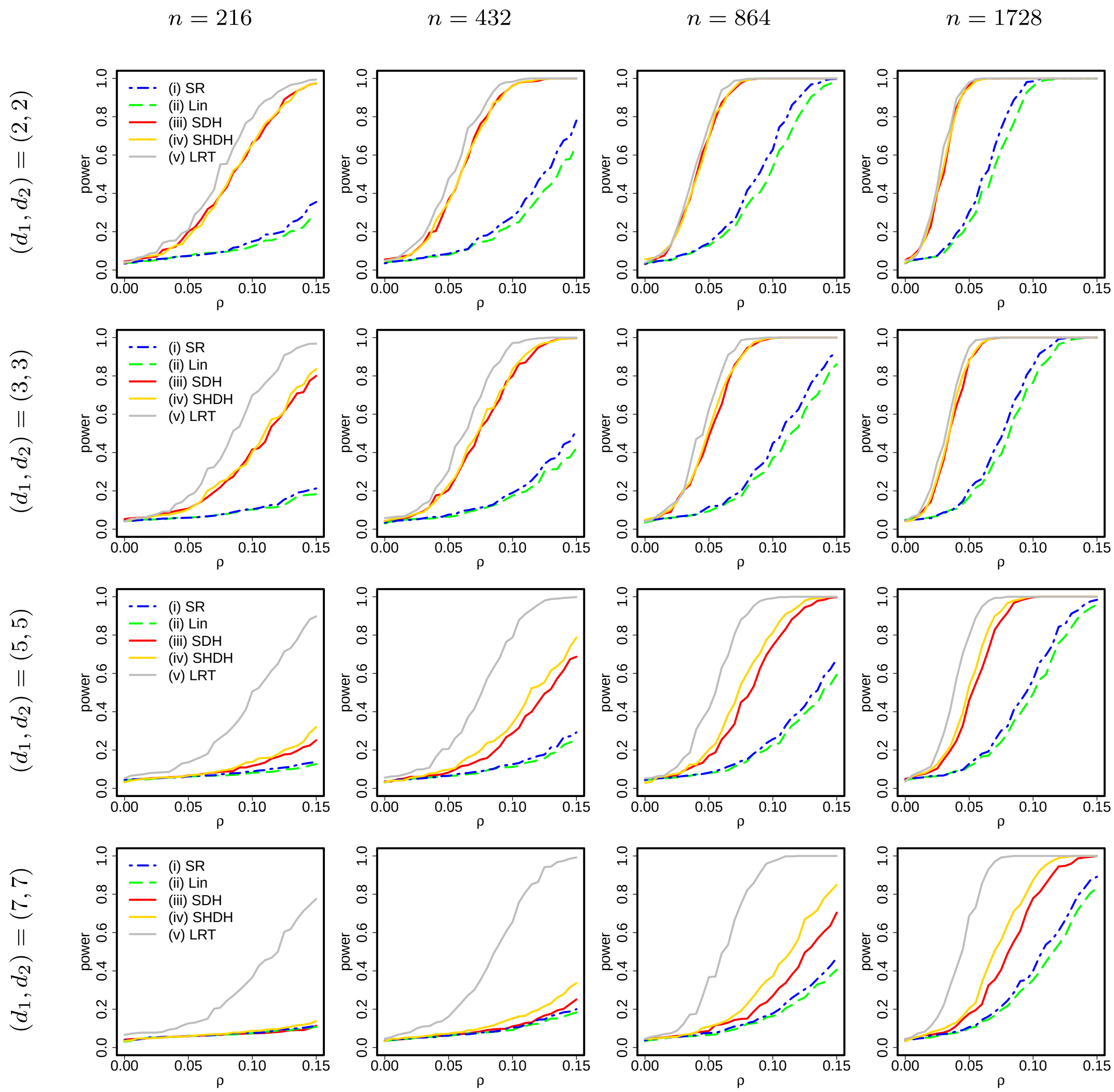


Figure 3.3: Empirical powers of the five competing tests in Example 3.5.3(c) ($\tau = 0.9$, high within-group correlation). The y-axis represents rejection frequencies based on 1,000 replicates, the x-axis represents ρ (the between-group correlation), and the blue, green, red, and gold lines represent the performance of (i) Szekely and Rizzo's original distance covariance test, (ii) Lin's marginal rank version of the distance covariance test, (iii) Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test, (iv) the center-outward normal-score version of the distance covariance test, and (v) the likelihood ratio test, respectively.

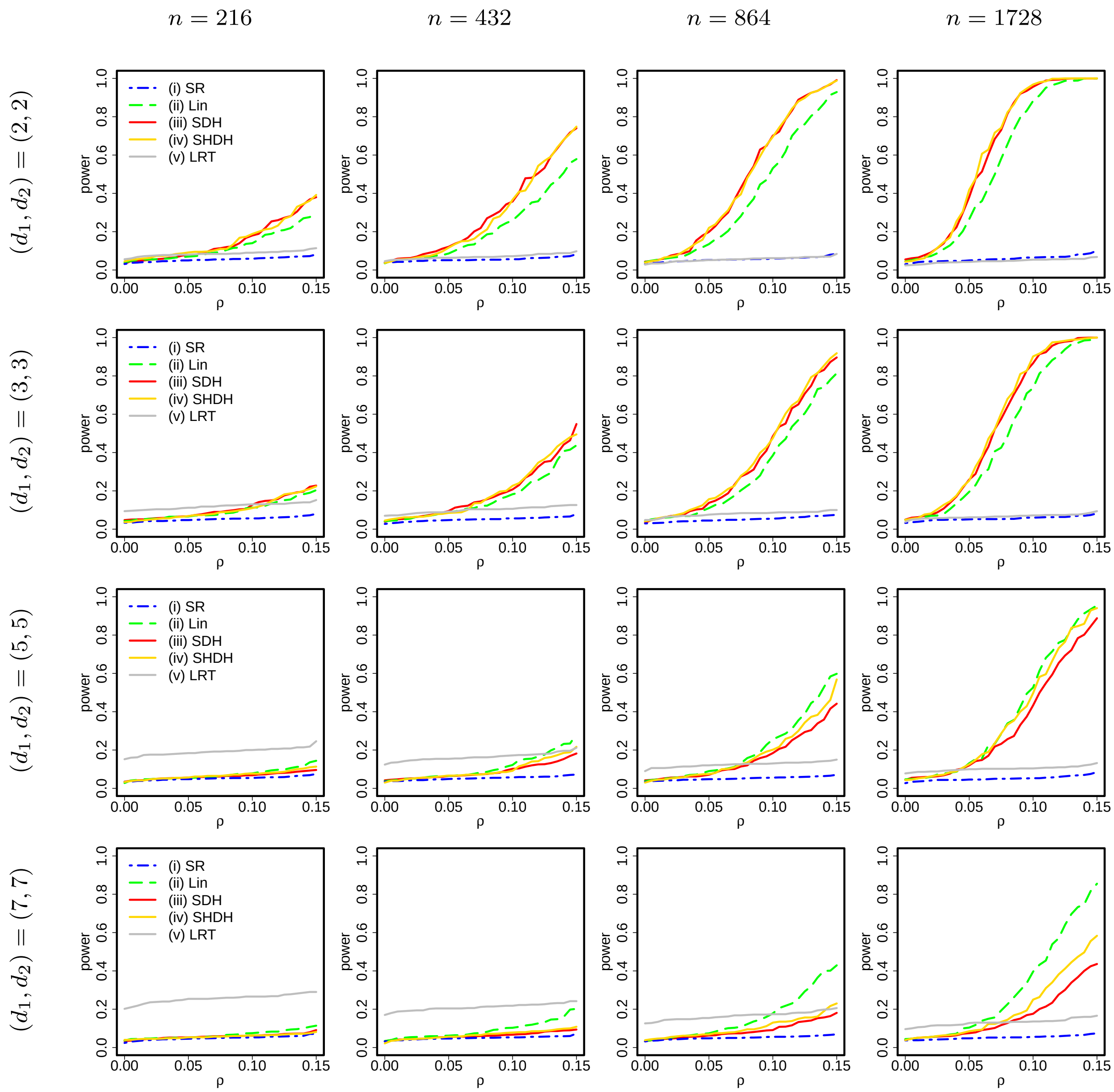


Figure 3.4: Empirical powers of the five competing tests in Example 3.5.4. The y-axis represents rejection frequencies based on 1,000 replicates, the x-axis represents ρ (the between-group correlation), and the blue, green, red, and gold lines represent the performance of (i) Szekely and Rizzo's original distance covariance test, (ii) Lin's marginal rank version of the distance covariance test, (iii) Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test, (iv) the center-outward normal-score version of the distance covariance test, and (v) the likelihood ratio test, respectively.

Chapter 4

ON THE POWER OF CHATTERJEE'S RANK CORRELATION

4.1 Introduction

Let X_1, X_2 be two real-valued random variables defined on a common probability space. We will be concerned with testing the null hypothesis

$$H_0 : X_1 \text{ and } X_2 \text{ are independent}, \tag{4.1.1}$$

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, e.g., methods for causal discovery (Maathuis et al., 2019, Section 18.6.3).

The problem of testing independence has been examined from a number of different perspectives; see, for example, the work of Meynaoui et al. (2019), Berrett et al. (2021), and Kim et al. (2020a), 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), but also 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 fact 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 and Dassios (2014) proposed a new test of independence and showed its consistency for bivariate distributions that are discrete, absolutely continuous, or a mixture of both 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 considers the correlation of Bergsma and 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 non-normal 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 is also able to detect perfect functional dependence (see also Gamboa et al., 2018) and in a recent paper that received much attention, Chatterjee (2021) gives a very simple rank correlation, with no tuning parameter involved, that surprisingly estimates ξ and has an asymptotically normal null distribution.

This paper compares Chatterjee's and also Dette–Siburg–Stoimenov's rank correlation coefficients to the three obvious competitors given by the D of Hoeffding (1948), the R of Blum et al. (1961), and the τ^* of Bergsma and 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 in a family of distributions $\mathcal{F}$ if for all pairs (X_1, X_2) with joint distribution in $\mathcal{F}$, it holds that $\mu(X_1, X_2) = 0$ 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 non-linear, non-monotone relationship, 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.* Computing 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 Section 4.2, recent advances clarify that D, R, and τ^* can be computed similarly efficiently.

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

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 then invoking a type of permutation central limit theorem that was established in Chao et al. (1993). We found that a direct use of this technique to analyse the local power is hard. In recent related work we were able to overcome a similar issue in a related multivariate setting (Shi et al., 2021a; Deb and Sen, 2021) by developing a suitable Hájek representation theory (Shi et al., 2020). Applying this philosophy here, we build a particular form of the projected statistic that was 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. Chapter 2.9 in van der Vaart and Wellner, 1996), we are then able to show that the test based on Chatterjee's rank correlation coefficient is in fact *rate sub-optimal* against the two considered local alternative families; recall point (iii) above. Our theoretical analysis thus echos Chatterjee's empirical observation, that is, his test of independence can suffer from low power; see Remark 4.3.4 below. In contrast, the tests based on the more established coefficients D, R, and τ^* are all rate-optimal for all considered local alternative families. We therefore consider the latter 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 non-trivial power against certain alternatives in finite-sample simulations. A theoretical study of this phenomenon, however, has to be left to the future due to involved technical difficulties. The proofs of our claims, including details on examples, are given in the supplementary material.

As we were completing the manuscript, we became aware of independent work by Cao and Bickel (2020), who accomplished a similar local power analysis for Chatterjee's correlation coefficient and presented a result that is similar to our Theorem 4.3.1, Claim (4.3.5). 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 and Bickel concentrates on correlation measures that are 1 if and only if one variable is a shape-restricted function of the other variable, while our interest is in comparing consistent tests of independence.

4.2 Rank correlations and independence tests

4.2.1 Considered rank correlations 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 considered pair of random variables (X_1, X_2), and F_1 and F_2 are the respective marginal distribution functions. Throughout, $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ is a sample comprised of n independent copies of (X_1, X_2).

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

$$r_{[i]} := \sum_{j=1}^{n} \mathbb{1}\left(X_{2[j]} \leq X_{2[i]}\right) \tag{4.2.1}$$

with $I(\cdot)$ representing the indicator function, and $\ell_{[i]} := \sum_{j=1}^{n} \mathbb{1}(X_{2[j]} \geq X_{2[i]})$. We emphasize that if F_2 is continuous, then there are almost surely no ties among $X_{21}, \ldots, X_{2n}$, in which case $r_{[i]}$ is simply the rank of $X_{2[i]}$ among $X_{2[1]}, \ldots, X_{2[n]}$.

Definition 4.2.1. The correlation coefficient of Chatterjee (2021) is

$$\xi_n := 1 - \frac{n \sum_{i=1}^{n-1} |r_{[i+1]} - r_{[i]}|}{2 \sum_{i=1}^{n} \ell_{[i]}(n - \ell_{[i]})}. \tag{4.2.2}$$

If there are no ties among $X_{21}, \ldots, X_{2n}$, it holds that

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

Chatterjee (2021) proved that ξ_n estimates the correlation measure

$$\xi := \frac{\int \mathrm{Var}[\mathrm{E}\{\mathbb{1}(X_2 \geq x) \mid X_1\}] \mathrm{d}F_2(x)}{\int \mathrm{Var}\{\mathbb{1}(X_2 \geq x)\} \mathrm{d}F_2(x)}.$$

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

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

Definition 4.2.2. Let K be a symmetric and twice continuously differentiable kernel with compact support, and let $\overline{K}(x) := \int_{-\infty}^{x} K(t)\mathrm{d}t$. Let $h_1, h_2 > 0$ be bandwidths that are chosen such that they tend to zero with

$$nh_1^3 \to \infty, \quad nh_1^4 \to 0, \quad nh_2^4 \to 0, \quad nh_1h_2 \to \infty \tag{4.2.3}$$

as $n \to \infty$. Define

$$\zeta_n(u_1, u_2) := \frac{1}{nh_1} \sum_{i=1}^{n} K\Big(\frac{u_1 - i/n}{h_1}\Big)\overline{K}\Big(\frac{u_2 - r_{[i]}/n}{h_2}\Big) \tag{4.2.4}$$

with $r_{[i]}$ as in (4.2.1). Then the Dette–Siburg–Stoimenov's correlation coefficient is

$$\xi_n^* := 6\int_0^1\int_0^1 \{\zeta_n(u_1, u_2)\}^2 \mathrm{d}u_1\mathrm{d}u_2 - 2.$$

Next we introduce two classical rank correlations of Hoeffding (1948) and Blum et al. (1961), both of which 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 4.2.3. Hoeffding's correlation measure is defined as

$$D := \int \Big\{F(x_1, x_2) - F_1(x_1)F_2(x_2)\Big\}^2 \mathrm{d}F(x_1, x_2).$$

It is unbiasedly estimated by the correlation coefficient

$$D_n := \frac{1}{n(n-1)\cdots(n-4)} \sum_{i_1 \neq \ldots \neq i_5} \frac{1}{4}$$
$$\Big[\Big\{\mathbb{1}\Big(X_{1i_1} \le X_{1i_5}\Big) - \mathbb{1}\Big(X_{1i_2} \le X_{1i_5}\Big)\Big\}\Big\{\mathbb{1}\Big(X_{1i_3} \le X_{1i_5}\Big) - \mathbb{1}\Big(X_{1i_4} \le X_{1i_5}\Big)\Big\}\Big]$$
$$\Big[\Big\{\mathbb{1}\Big(X_{2i_1} \le X_{2i_5}\Big) - \mathbb{1}\Big(X_{2i_2} \le X_{2i_5}\Big)\Big\}\Big\{\mathbb{1}\Big(X_{2i_3} \le X_{2i_5}\Big) - \mathbb{1}\Big(X_{2i_4} \le X_{2i_5}\Big)\Big\}\Big], \tag{4.2.5}$$

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

Definition 4.2.4. Blum–Kiefer–Rosenblatt's correlation measure is defined as

$$R := \int \Big\{F(x_1,x_2) - F_1(x_1)F_2(x_2)\Big\}^2 \mathrm{d}F_1(x_1)\mathrm{d}F_2(x_2).$$

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

$$\begin{aligned} R_n := \frac{1}{n(n-1)\cdots(n-5)} \sum_{i_1\neq\ldots\neq i_6} \frac{1}{4} \\ \Big[\Big\{\mathbb{1}\Big(X_{1i_1} \le X_{1i_5}\Big) - \mathbb{1}\Big(X_{1i_2} \le X_{1i_5}\Big)\Big\}\Big\{\mathbb{1}\Big(X_{1i_3} \le X_{1i_5}\Big) - \mathbb{1}\Big(X_{1i_4} \le X_{1i_5}\Big)\Big\}\Big] \\ \Big[\Big\{\mathbb{1}\Big(X_{2i_1} \le X_{2i_6}\Big) - \mathbb{1}\Big(X_{2i_2} \le X_{2i_6}\Big)\Big\}\Big\{\mathbb{1}\Big(X_{2i_3} \le X_{2i_6}\Big) - \mathbb{1}\Big(X_{2i_4} \le X_{2i_6}\Big)\Big\}\Big], \end{aligned} \tag{4.2.6}$$

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

More recently, Bergsma and Dassios (2014) introduced the following rank correlation, which is connected to work by Yanagimoto (1970). We refer the reader to Bergsma and Dassios (2014) for a motivation via con-/disconcordance of 4-point patterns and connections to Kendall's tau.

Definition 4.2.5. Write $\mathbb{1}(x_1, x_2 < x_3, x_4) := \mathbb{1}(\max\{x_1, x_2\} < \min\{x_3, x_4\})$. The Bergsma–Dassios–Yanagimoto's correlation measure is

$$\begin{aligned} \tau^* := {} & 4\mathrm{P}\Big(X_{11}, X_{13} < X_{12}, X_{14}\ ,\ X_{21}, X_{23} < X_{22}, X_{24}\Big) \\ & + 4\mathrm{P}\Big(X_{11}, X_{13} < X_{12}, X_{14}\ ,\ X_{22}, X_{24} < X_{21}, X_{23}\Big) \\ & - 8\mathrm{P}\Big(X_{11}, X_{13} < X_{12}, X_{14}\ ,\ X_{21}, X_{24} < X_{22}, X_{23}\Big). \end{aligned}$$

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

$$\tau_n^* := \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1\neq\ldots\neq i_4}$$

$$\Big\{\mathbb{1}\Big(X_{1i_1}, X_{1i_3} < X_{1i_2}, X_{1i_4}\Big) + \mathbb{1}\Big(X_{1i_2}, X_{1i_4} < X_{1i_1}, X_{1i_3}\Big)$$
$$- \mathbb{1}\Big(X_{1i_1}, X_{1i_4} < X_{1i_2}, X_{1i_3}\Big) - \mathbb{1}\Big(X_{1i_2}, X_{1i_3} < X_{1i_1}, X_{1i_4}\Big)\Big\}$$
$$\Big\{\mathbb{1}\Big(X_{2i_1}, X_{2i_3} < X_{2i_2}, X_{2i_4}\Big) + \mathbb{1}\Big(X_{2i_2}, X_{2i_4} < X_{2i_1}, X_{2i_3}\Big)$$
$$- \mathbb{1}\Big(X_{2i_1}, X_{2i_4} < X_{2i_2}, X_{2i_3}\Big) - \mathbb{1}\Big(X_{2i_2}, X_{2i_3} < X_{2i_1}, X_{2i_4}\Big)\Big\}. \tag{4.2.7}$$

Remark 4.2.1 (Relation between D_n, R_n, and τ_n^*)**.** As conveyed by Equation (6.1) in Drton et al. (2020), as long as $n \geq 6$ and there are no ties in the data, it holds that $12D_n + 24R_n = \tau_n^*$. Consequently, $12D + 24R = \tau^*$ given continuity but not necessarily absolute continuity of F; compare page 62 of Yanagimoto (1970).

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

Proposition 4.2.1 (Computational efficiency)**.** *If data have no ties, then ξ_n, D_n, R_n, and τ_n^* can all be computed in $O(n \log n)$ time.*

Proof. It is evident from its simple form 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, Section 5); see also Weihs et al. (2018, page 557). The claim about τ_n^* is based on recent new methods due to Even-Zohar and Leng (2021, Corollary 1.3) 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 4.2.1. □

Remark 4.2.2 (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 considered kernel and the bandwidth choice. Chatterjee (2021, Remark 5) suggests a time complexity of $O(n^{5/3})$. Indeed, for a symmetric and four times continuously differentiable kernel K that has compact support, there is a choice of bandwidths h_1, h_2 that satisfies the requirements of Definition 4.2.2 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 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 $O(n^{-1/2-\epsilon}) = o(n^{-1/2})$. Due to 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 $\epsilon \leq 1/48$.

Remark 4.2.3 (Computation with ties)**.** When the data can be considered as 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 due to 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 from (4.2.2), 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 $O(n \log n)$ but, as we shall emphasize in Section 4.4, 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, Sec. 3) or the slighly more memory intensive but faster $O(n^2)$ algorithm of Heller and Heller (2016b, Sec. 2.2). Computation of R_n with ties is also $O(n^2)$ (Weihs et al., 2018; Weihs, 2019).

4.2.2 Consistency

In the rest of this section as well as in Section 4.3, we will always assume that the joint distribution function F is continuous, though not necessarily jointly absolutely continuous with regard to the Lebesgue measure. Accordingly, both $X_{11}, \ldots, X_{1n}$ and $X_{21}, \ldots, X_{2n}$ are free of ties with probability one. To clearly state the following results, we introduce three families of bivariate distributions specified via their joint distribution function F:

$$\begin{aligned}\mathcal{F}^{\mathbf{c}} &:= \big\{F : F \text{ is continuous as a bivariate function}\big\},\\ \mathcal{F}^{\text{ac}} &:= \big\{F : F \text{ is absolutely continuous with regard to the Lebesgue measure}\big\},\\ \mathcal{F}^{\text{DSS}} &:= \big\{F \in \mathcal{F}^{\mathbf{c}} : F \text{ has a copula } C(u_1, u_2) \text{ that is three and two times continuously}\\ &\qquad \text{differentiable with respect to the arguments } u_1 \text{ and } u_2\text{, respectively}\big\}. \qquad (4.2.8)\end{aligned}$$

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 $\xrightarrow{\mathsf{p}}$.

Proposition 4.2.2 (Consistency of estimators)**.** *For any $F \in \mathcal{F}^{\mathbf{c}}$ and $n \to \infty$, we have*

$$\xi_n \xrightarrow{\mathsf{p}} \xi, \quad D_n \xrightarrow{\mathsf{p}} D, \quad R_n \xrightarrow{\mathsf{p}} R, \quad \textit{and} \quad \tau_n^* \xrightarrow{\mathsf{p}} \tau^*.$$

If in addition $F \in \mathcal{F}^{\text{DSS}}$ and K, h_1, h_2 satisfy all assumptions stated in Definition 4.2.2, then also $\xi_n^* \xrightarrow{\mathsf{p}} \xi$.

Proof. The claim about ξ_n is Theorem 1.1 in Chatterjee (2021), and the one about ξ_n^* is proved in the supplement Section C.1.1 based on a revised version of Theorem 3 in Dette et al. (2013). The remaining claims are immediate from U-statistics theory (e.g., Proposition 1 in Weihs et al., 2018, 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 $F \in \mathcal{F}^{\boldsymbol{c}}$; this follows from Remark 4.2.1. The consistency properties for continuous observations can be summarized as follows.

Proposition 4.2.3 (Consistency of correlation measures)**.** *Each one of the correlation measures ξ, R, and τ^* is consistent for the entire class $\mathcal{F}^{\boldsymbol{c}}$, that is, if $F \in \mathcal{F}^{\boldsymbol{c}}$, 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 $\mathcal{F}^{\boldsymbol{c}}$.*

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 p. 490 in Blum et al. (1961). The consistency of τ^* was established for $\mathcal{F}^{\mathrm{ac}}$ in Theorem 1 in Bergsma and Dassios (2014), and that for $\mathcal{F}^{\boldsymbol{c}}$ can be shown via Remark 4.2.1; compare 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). □

4.2.3 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. We use $\rightsquigarrow$ to denote convergence in distribution.

Proposition 4.2.4 (Limiting null distributions)**.** *Suppose $F \in \mathcal{F}^{\boldsymbol{c}}$ has X_1 and X_2 independent. As $n \to \infty$, it holds that*

(i) for Chatterjee's correlation coefficient ξ_n, $n^{1/2}\xi_n \rightsquigarrow N(0, 2/5)$ (Theorem 2.1 in Chatterjee, 2021);

(ii) for Dette–Siburg–Stoimenov's correlation coefficient ξ_n^*, $n^{1/2}\xi_n^* \xrightarrow{\mathsf{p}} 0$ assuming that $F \in \mathcal{F}^{\mathrm{DSS}}$ and K, h_1, h_2 satisfy all assumptions stated in Definition 4.2.2 (revised

version of Theorem 3 in Dette et al., 2013; see Section C.1.2 of the supplementary material);

(iii) for $\mu \in \{D, R, \tau^*\}$,

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

where

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

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

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

$$T^{\xi_n}_{\alpha} := \mathbb{1}\big\{n^{1/2}|\xi_n| > (2/5)^{1/2} \cdot z_{1-\alpha/2}\big\}.$$

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

$$T^{\mu_n}_{\alpha} := \mathbb{1}\big(n\,\mu_n > q^{\mu}_{1-\alpha}\big), \quad q^{\mu}_{1-\alpha} := \inf\Big[x : \mathrm{P}\Big\{\sum_{v_1,v_2=1}^{\infty} \lambda^{\mu}_{v_1,v_2}\Big(\xi^2_{v_1,v_2} - 1\Big) \le x\Big\} \ge 1 - \alpha\Big],$$

where $\lambda^{\mu}_{v_1,v_2}$ and ξ_{v_1,v_2}, $v_1, v_2 = 1, \ldots, n, \ldots$ were presented in Proposition 4.2.4. We note that Weihs (2019) gives 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 non-degenerate limiting null distribution is currently unknown.

Given the distribution-freeness of ranks for the class $\mathcal{F}^{\mathbf{c}}$, Proposition 4.2.4 yields uniform asymptotic validity of the tests just defined. Moreover, Propositions 4.2.2–4.2.3 yield consistency at fixed alternatives. We summarize these facts below.

Proposition 4.2.5 (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*

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}^{\mathbf{c}}} \mathrm{P}(T_\alpha^{\mu_n} = 1 \mid H_0) = \alpha. \tag{4.2.9}$$

Moreover, these tests are consistent, i.e., for fixed $F \in \mathcal{F}^{\mathbf{c}}$ *such that* X_1 *and* X_2 *are dependent and* $\mu_n \in \{\xi_n, R_n, \tau_n^*\}$*, it holds that*

$$\lim_{n\to\infty} \mathrm{P}(T_\alpha^{\mu_n} = 1 \mid H_1) = 1. \tag{4.2.10}$$

The conclusion (4.2.10) *holds for* $\mu_n = D_n$ *if assuming further that* $F \in \mathcal{F}^{\mathrm{ac}}$*.*

4.3 Local power analysis

This section investigates the local power of the four rank correlation-based tests of H_0 introduced in Section 4.2.3. To this end, we consider two classical and well-used families of alternatives to the null hypothesis of independence: *rotation alternatives* (*Konijn alternatives*; Konijn, 1956) and *mixture alternatives* (*Farlie-type alternatives*; Farlie, 1960, 1961; see also Dhar et al., 2016).

(A) 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

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} := \begin{pmatrix} 1 & \Delta \\ \Delta & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{A}_\Delta \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{A}_\Delta \boldsymbol{Y}. \tag{4.3.1}$$

For all $\Delta \in (-1, 1)$, the matrix $\mathbf{A}_\Delta$ is clearly full rank and invertible. For any $\Delta \in (-1, 1)$, let $f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta)$ denote the density of $\boldsymbol{X} = \mathbf{A}_\Delta \boldsymbol{Y}$. We then make the following assumptions on Y_1, Y_2.

Assumption 4.3.1. *It holds that*

(i) the distributions of $\boldsymbol{X}$ *have a common support for all* $\Delta \in (-1, 1)$, *so that without loss of generality* $\mathcal{X} := \{\boldsymbol{x} : f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta) > 0\}$ *is independent of* Δ;

(ii) the density f_k *is absolutely continuous with non-constant logarithmic derivative* $\rho_k := f_k'/f_k$, $k = 1, 2$;

(iii) the Fisher information of $\boldsymbol{X}$ *relative to* Δ *at the point* 0, *denoted* $\mathcal{I}_{\boldsymbol{X}}(0)$, *is strictly positive, and* $\mathrm{E}\{(Y_k)^2\} < \infty$, $\mathrm{E}[\{\rho_k(Y_k)\}^2] < \infty$ *for* $k = 1, 2$.

Remark 4.3.1. Assumption 4.3.1(ii),(iii) implies $\mathrm{E}\{\rho_k(Y_k)\} = 0$ and $\mathcal{I}_{\boldsymbol{X}}(0) < \infty$.

Example 4.3.1. Suppose $f_k(z)$ is absolutely continuous and positive for all real numbers z, $k = 1, 2$. If

$$\mathrm{E}\big(Y_k\big) = 0, \quad \mathrm{E}\big\{\big(Y_k\big)^2\big\} < \infty, \quad \mathrm{E}\big[\big\{\rho_k\big(Y_k\big)\big\}^2\big] < \infty, \quad \text{for } k = 1, 2, \tag{4.3.2}$$

then Assumption 4.3.1 holds. As a special case, Assumption 4.3.1 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 two.

(B) Mixture alternatives. Consider the following mixture alternatives that were used in Dhar et al. (2016, Sec. 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 that 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 $\boldsymbol{X} = (X_1, X_2)$:

$$F_{\boldsymbol{X}} := (1 - \Delta)F_0 + \Delta G, \tag{4.3.3}$$

with $0 \leq \Delta \leq 1$.

We make the following additional assumptions on F_0 and G.

Assumption 4.3.2. *It holds that*

(i) the distribution G is absolutely continuous with respect to F_0 and $s(\boldsymbol{x}) := g(\boldsymbol{x})/f_0(\boldsymbol{x}) - 1$ is continuous;

(ii) the conditional expectation $\mathrm{E}\{s(\boldsymbol{Y})|Y_1\} = 0$ almost surely for $\boldsymbol{Y} = (Y_1, Y_2) \sim F_0$;

(iii) the function s is not additively separable, i.e., there do not exist univariate functions h_1 and h_2 such that $s(\boldsymbol{x}) = h_1(x_1) + h_2(x_2)$;

(iv) the Fisher information $\mathcal{I}_{\boldsymbol{X}}(0) > 0$.

Remark 4.3.2. In this model, $g(\boldsymbol{x})/f_0(\boldsymbol{x})$ is continuous and has compact support, which guarantees that $\mathcal{I}_{\boldsymbol{X}}(0) < \infty$.

Example 4.3.2. (Farlie alternatives) Let G in (4.3.3) be given as

$$G(x_1, x_2) = F_1(x_1)F_2(x_2)\Big[1 + \big\{1 - F_1(x_1)\big\}\big\{1 - F_2(x_2)\big\}\Big].$$

Then Assumption 4.3.2 is satisfied (Morgenstern, 1956; Gumbel, 1958; Farlie, 1960). Notice also that $\mathrm{E}\{s(\boldsymbol{Y})|Y_2\} = 0$ almost surely for $\boldsymbol{Y} = (Y_1, Y_2) \sim F_0$.

Example 4.3.3. Let the density f_2 be symmetric around 0, and consider two univariate functions h_1 and h_2 that are both non-constant 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 $s(\boldsymbol{x}) = h_1(x_1)h_2(x_2)$ and then Assumption 4.3.2 holds. For example, we can take $f_1(t) = f_2(t) = 1/2 \times \mathbb{1}(-1 \leq t \leq 1)$, $h_1(t) = |1 - 2\Psi(t)|$, and $h_2(t) = 1 - 2\Psi(t)$, where Ψ denotes the distribution function of the uniform distribution on $[-1, 1]$. In this case, $\mathrm{E}\{s(\boldsymbol{Y})|Y_2\}$ is not almost surely zero for $\boldsymbol{Y} = (Y_1, Y_2) \sim F_0$.

For a local power analysis in any one of the two considered alternative families, we examine the asymptotic power along a respective sequence of alternatives obtained as

$$H_{1,n}(\Delta_0) : \Delta = \Delta_n, \text{ where } \Delta_n := n^{-1/2}\Delta_0 \tag{4.3.4}$$

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

Theorem 4.3.1 (Power analysis)**.** *Suppose the considered sequences of local alternatives are formed such that Assumption 4.3.1 or 4.3.2 holds when considering a family of type (A) or (B), respectively. Then concerning any sequence of alternatives given in* (4.3.4)*,*

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

$$\lim_{n\to\infty} \mathrm{P}\{T_\alpha^{\xi_n} = 1 \mid H_{1,n}(\Delta_0)\} = \alpha; \tag{4.3.5}$$

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

$$\lim_{n\to\infty} \mathrm{P}\{T_\alpha^{\mu_n} = 1 \mid H_{1,n}(\Delta_0)\} \geq 1 - \beta, \tag{4.3.6}$$

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

In contrast to Theorem 4.3.1, Proposition 4.3.1 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 (4.3.5) and (4.3.6) manifests that the size-α tests based on one of D_n, R_n, τ_n^* are rate-optimal against the considered local alternatives, 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 sub-optimal.

Proposition 4.3.1 (Rate-optimality)**.** *Concerning any one of the two local alternative families and any sequence of alternatives given in* (4.3.4)*, as long as the corresponding Assump-*

tion 4.3.1 or 4.3.2 holds, we have that for any number $\beta > 0$ satisfying $\alpha + \beta < 1$ there exists a constant $c_\beta > 0$ only depending on β such that

$$\inf_{\overline{T}_\alpha \in \mathcal{T}_\alpha} \mathrm{P}\{\overline{T}_\alpha = 0 \mid H_{1,n}(c_\beta)\} \geq 1 - \alpha - \beta$$

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

Remark 4.3.3. Assumptions 4.3.1 and 4.3.2 are technical conditions imposed to ensure that (i) the two considered sequences of alternatives are all locally asymptotically normal (van der Vaart, 1998, Chapter 7), i.e., the log likelihood 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 for a use of the conditional multiplier central limit theorem (cf. Chapter 2.9 in van der Vaart and Wellner, 1996) that appears to be the key in 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 local power of independence tests.

Remark 4.3.4. We note that the linear, step function, W-shaped, sinusoid, and circular alternatives considered in Chatterjee (2021, Section 4.3) can all be viewed as generalized rotation alternatives. The proof techniques used in this paper are hence directly applicable to these five alternatives by means of a re-parametrization. To illustrate this point, consider, for example, the following alternative motivated by Chatterjee (2021, Section 4.3):

$$X_1 = Y_1 \quad \text{and} \quad X_2 = \Delta g(Y_1) + Y_2, \tag{4.3.7}$$

where Y_1 and Y_2 are independent and absolutely continuous with respective densities f_1, f_2. Notice that model (4.3.7) and the one used in Chatterjee (2021, Section 4.3) are equivalent for rank-based tests as ranks are scale invariant. Assume then that

(i) the distributions of $\boldsymbol{X} = (X_1, X_2)$ have a common support for all $\Delta \in (-1, 1)$;

(ii) the density f_2 is absolutely continuous with non-constant logarithmic derivative $\rho_2 := f_2'/f_2$ with $0 < \mathrm{E}[\{\rho_2(Y_2)\}^2] < \infty$;

(iii) the function g is non-constant and measurable such that $0 < \mathrm{E}[\{g(Y_1)\}^2] < \infty$.

Claims (4.3.5) and (4.3.6) will then hold for the alternatives (4.3.7) in observation of arguments similar to those made in the proof of Theorem 4.3.1 for the rotation alternatives (A).

Remark 4.3.5. Cao and Bickel (2020, Section 4.4) performed a local power analysis for Chatterjee's ξ_n under a set of assumptions that differs 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 4.2.1, 4.2.3–4.2.5. To this end, we focused on rotation and mixture alternatives from the literature. However, from the proof techniques in Section C.1.8 of the supplementary material, it is evident that Claims (4.3.5) and (4.3.6) hold for further types of local alternative families. For the former claim, which concerns lack of power of Chatterjee's ξ_n, this point has been pursued in Section 4.4 of Cao and Bickel (2020).

4.4 Rank correlations for discontinuous distributions

In this section, we drop the continuity assumption of F made in Sections 4.2–4.3, and allow for ties to exist with a nonzero probability. Among the five correlation coefficients, ξ_n^* is no longer an appropriate estimator when F is not continuous. We will only discuss the properties of the other four estimators ξ_n, D_n, R_n, and τ_n^*.

Recall that the computation issue has been address in Remark 4.2.3. Our first result in this section focuses on approximation consistency of the correlation coefficients ξ_n, D_n, R_n and τ_n^* to their population quantities. To this end, we define the families of distribution more general than the ones considered so far as follows:

$$\mathcal{F} := \big\{F : F \text{ is a bivariate distribution function}\big\},$$
$$\mathcal{F}^* := \big\{F : F_2 \text{ is not degenerate, i.e., } F_2(x) \neq I(x \geq x_0) \text{ for any real number } x_0\big\},$$
$$\mathcal{F}^{\tau^*} := \big\{F : F \text{ is discrete, continuous, or a mixture of discrete and jointly absolutely continuous distribution functions}\big\}. \quad (4.4.1)$$

For the estimators ξ_n, D_n, R_n, and τ_n^*, the following result on consistency can be given.

Proposition 4.4.1 (Consistency of estimators)**.** *As $n \to \infty$, we have*

(i) for $F \in \mathcal{F}^*$, ξ_n converges in probability to ξ (Theorem 1.1 in Chatterjee, 2021);

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

The following proposition is a generalization of Proposition 4.2.3.

Proposition 4.4.2 (Consistency of correlation measures)**.** *The following are true:*

(i) for $F \in \mathcal{F}^*$, $\xi \geq 0$ with equality if and only if the pair is independent (Theorem 1.1 in Chatterjee, 2021);

(ii) for $F \in \mathcal{F}$, $D \geq 0$; for $F \in \mathcal{F}^{\mathrm{ac}}$, $D = 0$ if and only if the pair is independent (Theorem 3.1 in Hoeffding, 1948, Proposition 3 in Yanagimoto, 1970);

(iii) for $F \in \mathcal{F}$, $R \geq 0$ with equality if and only if the pair is independent (page 490 of Blum et al., 1961);

(iv) for $F \in \mathcal{F}^{\tau^*}$, $\tau^* \geq 0$ where equality holds if and only if the variables are independent (Theorem 1 in Bergsma and Dassios, 2014, Theorem 6.1 in Drton et al., 2020).

The asymptotic distribution theory from Section 4.2.3 can also be extended. As the continuity requirement is dropped, the central limit theorems 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 (Theorem 2.2 in Chatterjee, 2021). 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 Theorem 4.5 and Corollary 4.1 in Nandy et al. (2016) for further discussion. As a result, permutation analysis, which is unfortunately computationally much more intensive, is typically invoked to implement a test outside the realm of continuous distributions.

4.5 Simulation results

In order to further examine the power of the tests, we simulate data as a sample comprised 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 4.5.1. For the distribution of (X_1, X_2) we choose the 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 (4.3.1), where Y_1, Y_2 are both standard Gaussian and $\Delta_0 = 2$. This is an instance of our Example 4.3.1.

(b) The pair (X_1, X_2) is given by the mixture alternative (4.3.3), where

$$
\begin{aligned}
F_0\big(x_1, x_2\big) &= \Psi\big(x_1\big)\Psi\big(x_2\big), \\
G\big(x_1, x_2\big) &= \Psi\big(x_1\big)\Psi\big(x_2\big)\big[1 + \big\{1 - \Psi\big(x_1\big)\big\}\big\{1 - \Psi\big(x_2\big)\big\}\big],
\end{aligned}
$$

$\Psi(\cdot)$ denotes the distribution function of the uniform distribution on $[-1, 1]$, and $\Delta_0 = 10$. This is in accordance with our Example 4.3.2.

(c) The pair (X_1, X_2) is given by the mixture alternative (4.3.3), where the density functions of F and G, denoted by f_0 and g, are given by

$$f_0(x_1, x_2) = \psi(x_1)\psi(x_2),$$
$$g(x_1, x_2) = \psi(x_1)\psi(x_2)\big[1 + \big|1 - 2\Psi(x_1)\big|\big\{1 - 2\Psi(x_2)\big\}\big],$$

$\psi(t) = 1/2 \times \mathbb{1}(-1 \leq t \leq 1)$, and $\Delta_0 = 20$. This is an instance of our Example 4.3.3.

(d) The pair (X_1, X_2) is given by the generalized rotation alternative (4.3.7), 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 $\Delta_0 = 3$.

(e) The pair (X_1, X_2) is given by (4.3.7), where Y_1 is uniformly distributed on $[-1, 1]$, Y_2 is standard Gaussian, $g(t) = |t + 0.5|\mathbb{1}(t < 0) + |t - 0.5|\mathbb{1}(t \geq 0)$, and $\Delta_0 = 60$.

(f) The pair (X_1, X_2) is given by (4.3.7), 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 4.3.1–4.3.3. The latter three are motivated by step function, W-shaped, and sinusoid settings in which Chatterjee's correlation coefficient performs well; see Chatterjee (2021, Section 4.3).

Our focus is on comparing the empirical performance of the five tests $T_\alpha^{\xi_n}$, $T_\alpha^{\xi_n^*}$, $T_\alpha^{D_n}$, $T_\alpha^{R_n}$, $T_\alpha^{\tau_n^*}$. The first four tests are conducted using the asymptotics from Proposition 4.2.4. The last test is implemented with bandwidths chosen as $h_1 = h_2 = n^{-3/10}$ following the suggestion in Section 6.1 of Dette et al. (2013) 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 $n \in \{500, 1000, 5000, 10000\}$. For each of the six settings and four sample sizes, we conduct 1000 simulations.

Before turning to statistical properties, we contrast the computation times for calculating the five considered rank correlation coefficients first. Table 4.1 shows times in the considered

Table 4.1: A comparison of computation time for all the five correlation statistics. The computation time here is the total time in seconds of 1000 replicates.

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

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, and those of D_n, R_n, τ_n^* are made using the functions `.calc.hoeffding()`, `.calc.refined()`, and `.calc.taustar()` from R package **`independence`** (Even-Zohar, 2020a), respectively. All experiments are conducted on a laptop with a 2.6 GHz Intel Core i5 processor and a 8 GB memory. One observes the clear computational advantages of ξ_n, D_n, R_n, and τ_n^* over Dette et al. (2013)'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 4.2 shows the empirical powers of the five tests. The results confirm our earlier theoretical claims on the powers of the different tests in the different models, that 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 non-trivial power against certain alternatives; see results for Example 4.5.1(e),(f) in Table 4.2.

Table 4.2: Empirical powers of the five competing tests in Example 4.5.1. The empirical powers here are based on 1000 replicates.

n	ξ_n	ξ_n^*	D_n	R_n	τ_n^*	ξ_n	ξ_n^*	D_n	R_n	τ_n^*
	Results for Example 4.5.1(a)					Results for Example 4.5.1(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 4.5.1(b)					Results for Example 4.5.1(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 4.5.1(c)					Results for Example 4.5.1(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

4.6 Discussion

In this paper we considered independence tests based on the five rank correlations from Definitions 4.2.1–4.2.5. As we surveyed in Section 4.2, recent advances lead to little difference in the efficiency of known algorithms to compute these correlation coefficients. For continuous distributions, i.e., data without ties, all correlations except for Dette–Siburg–Stoimenov's ξ_n^* can be computed in nearly linear time. Moreover, all but Hoeffding's D give consistent tests of independence for arbitrary continuous distributions; consistency of D can be established for all absolutely continuous distributions.

Our main new contribution is a local power analysis for continuous distributions that revealed interesting differences in the power of the tests. This analysis features subtle differences but the take-away message is that ξ_n is suboptimal for testing independence, whereas the more classical D_n, R_n, and τ_n^* are rate optimal in the considered setup. This said, ξ_n and ξ_n^* have very appealing properties that do not pertain to independence but rather detection of perfect functional dependence. We refer the reader to Dette et al. (2013) and Chatterjee (2021) as well as Cao and Bickel (2020).

We summarize the properties discussed in our paper in Table 4.3. When referring to independence tests in this table we assume continuous observations, i.e., $F \in \mathcal{F}^c$. Moreover, when discussing ξ_n^*, we assume additionally that the kernel K and bandwidths h_1, h_2 satisfy all assumptions stated in Definition 4.2.2. 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 (4.2.8) and (4.4.1) for the definitions of table entries. The fourth row concerns consistency of independence tests assuming $F \in \mathcal{F}^c$. Finally, we summarize the rate-optimality and rate sub-optimality of five independence tests under two local alternatives (A) and (B) considered in Section 4.3.

Table 4.3: Properties of the five rank correlation coefficients defined in Definitions 4.2.1–4.2.5.

			ξ_n	ξ_n^*	D_n	R_n	τ_n^*
(i)	Computational efficiency	$F \in \mathcal{F}^c$	$O(n \log n)$	$O(n^{5/3})$	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$
		$F \in \mathcal{F}$	$O(n \log n)$	——	$O(n \log n)$	$O(n^2)$	$O(n^2)$
(ii)	Consistency of correlation measures		$F \in \mathcal{F}^{*}$ [(a)]	$F \in \mathcal{F}^*$	$F \in \mathcal{F}^{\mathrm{ac}}$	$F \in \mathcal{F}$	$F \in \mathcal{F}^{\tau^*}$
(ii')	Consistency of independence tests		$F \in \mathcal{F}^c$	$F \in \mathcal{F}^{\mathrm{DSS}}$	$F \in \mathcal{F}^{\mathrm{ac}}$	$F \in \mathcal{F}^c$	$F \in \mathcal{F}^c$
(iii)	Statistical efficiency	(A)	rate sub-optimal	——	rate-optimal	rate-optimal	rate-optimal
		(B)	rate sub-optimal	——	rate-optimal	rate-optimal	rate-optimal

[(a)] Recall the definitions of bivariate distribution families in (4.2.8) and (4.4.1)

BIBLIOGRAPHY

Agarwal, P. K. and Sharathkumar, R. (2014). Approximation algorithms for bipartite matching with metric and geometric costs. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, pages 555–564. ACM, New York, NY.

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows in metric spaces and in the space of probability measures* (2nd ed.). Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., Hoboken, NJ.

Angus, J. E. (1995). A coupling proof of the asymptotic normality of the permutation oscillation. *Probab. Engrg. Inform. Sci.*, 9(4):615–621.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25.

Azadkia, M. and Chatterjee, S. (2021+). A simple measure of conditional dependence. *Ann. Statist.* (in press).

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.*, 37(6B):3822–3840.

Bao, Z. (2019). Tracy-Widom limit for Kendall's tau. *Ann. Statist.*, 47(6):3504–3532.

Bao, Z., Pan, G., and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.*, 17(88):1–32.

Barbour, A. D. and Eagleson, G. K. (1986). Random association of symmetric arrays. *Stochastic Anal. Appl.*, 4(3):239–281.

Barndorff-Nielsen, O. (1963). On the limit behaviour of extreme order statistics. *Ann. Math. Statist.*, 34:992–1002.

Beirlant, J., Buitendag, S., del Barrio, E., Hallin, M., and Kamper, F. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math. Econom.*, 95:79–100.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.*, 81(2):608–650.

Bentkus, V. and Götze, F. (1997). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Related Fields*, 109(3):367–416.

Bergsma, W. (2006). A new correlation coefficient, its orthogonal decomposition and associated tests of independence. Available at arXiv:math/0604627v1.

Bergsma, W. (2011). Nonparametric testing of conditional independence by means of the partial copula. Available at arXiv:1101.4607v1.

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

Berrett, T. B., Kontoyiannis, I., and Samworth, R. J. (2021+). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* (in press).

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

Bhattacharya, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 81(3):575–602.

Bhattacharya, R. N. and Ranga Rao, R. (1986). *Normal approximation and asymptotic expansions* (Rpt. ed.). Robert E. Krieger Publishing Co., Inc., Melbourne, FL.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.*, 21(4):593–600.

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

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(501):265–277.

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(3):1496–1525.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388.

Cao, S. and Bickel, P. J. (2020). Correlations with tailored extremal properties. Available at arXiv:2008.10177v2.

Chao, C.-C., Bai, Z., and Liang, W.-Q. (1993). Asymptotic normality for oscillation of permutation. *Probab. Engrg. Inform. Sci.*, 7(2):227–235.

Chatterjee, S. (2021+). A new coefficient of correlation. *J. Amer. Statist. Assoc.* (in press).

Chaudhuri, A. and Hu, W. (2019). A fast algorithm for computing distance correlation. *Comput. Statist. Data Anal.*, 135:15–24.

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.*, 29(4):972–994.

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

Deb, N., Bhattacharya, B. B., and Sen, B. (2021). Efficiency lower bounds for distribution-free Hotelling-type two-sample tests based on optimal transport. Available at arXiv:2104.01986v1.

Deb, N. and Sen, B. (2021+). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* (in press).

del Barrio, E., Cuesta-Albertos, J. A., Hallin, M., and Matrán, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions. Available at arXiv:1806.01238v1.

del Barrio, E., González-Sanz, A., and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.*, 180:104671, 13.

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

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

Dinic, E. A. and Kronrod, M. A. (1969). An algorithm for solving the assignment problem. *Dokl. Akad. Nauk SSSR*, 189(1):23–25.

Drton, M., Han, F., and Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.*, 48(6):3206–3227.

Dunford, N. and Schwartz, J. T. (1963). *Linear operators. Part II: Spectral theory. Self adjoint operators in Hilbert space*. With the assistance of William G. Bade and Robert G. Bartle. Interscience Publishers John Wiley & Sons New York-London.

Edmonds, J. and Karp, R. M. (1970). Theoretical improvements in algorithmic efficiency for network flow problems. In *Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969)*, pages 93–96. Gordon and Breach, New York.

Edmonds, J. and Karp, R. M. (1972). Theoretical improvements in algorithmic efficiency for network flow problems. *J. Assoc. Comput. Mach.*, 19(2):248–264.

Even-Zohar, C. (2020a). *independence: Fast Rank-Based Independence Testing*. R package version 1.0.1.

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

Even-Zohar, C. and Leng, C. (2021). Counting small permutation patterns. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2288–2302, Philadelphia, PA. Society for Industrial and Applied Mathematics (SIAM).

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47:307–323.

Farlie, D. J. G. (1961). The asymptotic efficiency of Daniels's generalized correlation coefficients. *J. Roy. Statist. Soc. Ser. B*, 23:128–142.

Feuerverger, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.*, 61(3):419–433.

Fukumizu, K., Bach, F. R., and Gretton, A. (2007). Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.*, 8(14):361–383.

Gabow, H. N. and Tarjan, R. E. (1989). Faster scaling algorithms for network problems. *SIAM J. Comput.*, 18(5):1013–1036.

Gamboa, F., Klein, T., and Lagnoux, A. (2018). Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA J. Uncertain. Quantif.*, 6(2):522–548.

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(3):677–693.

Gieser, P. W. (1993). *A new nonparametric test for independence between two sets of variates*. PhD thesis, University of Florida. Available at https://ufdc.ufl.edu/AA00003658/00001 and https://www.proquest.com/docview/304041219.

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

Gini, C. (1914). *L'ammontare e la composizione della ricchezza delle nazioni*, volume 62. Fratelli Bocca.

Götze, F. and Zaitsev, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *Ann. Probab.*, 42(1):354–397.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005a). Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic learning theory*, volume 3734 of

Lecture Notes in Comput. Sci., pages 63–77. Springer-Verlag Berlin Heidelberg, Berlin, Germany.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S. T., editors, *Advances in Neural Information Processing Systems 20*, pages 984–991. Curran Associates, Inc., Red Hook, NY.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b). Kernel methods for measuring independence. *J. Mach. Learn. Res.*, 6(Dec):2075–2129.

Gretton, A., Smola, A., Bousquet, O., Herbrich, R., Belitski, A., Augath, M., Murayama, Y., Pauls, J., Schölkopf, B., and Logothetis, N. (2005c). Kernel constrained covariance for dependence measurement. In Cowell, R. G. and Ghahramani, Z., editors, *AISTATS05*, pages 112–119. Society for Artificial Intelligence and Statistics.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric estimation under shape constraints: Estimators, algorithms and asymptotics*, volume 38 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York.

Gumbel, E. J. (1958). Distributions à plusieurs variables dont les marges sont données. *C. R. Acad. Sci. Paris*, 246:2717–2719.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York-London; Academia Publishing House of the Czechoslovak Academy of Sciences, Prague.

Hallin, M. (1994). On the Pitman non-admissibility of correlogram-based methods. *J. Time Ser. Anal.*, 15(6):607–611.

Hallin, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: a measure transportation approach. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

Hallin, M. (2021). Measure transportation and statistical decision theory. *Annu. Rev. Stat. Appl.* (in press).

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

Hallin, M., Hlubinka, D., and Hudecová, Š. (2020). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. Available at arXiv:2007.15496v1.

Hallin, M., La Vecchia, D., and Liu, H. (2021+b). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.* (in press).

Hallin, M., Mordant, G., and Segers, J. (2021c). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.*, 15(1):1328–1371.

Hallin, M. and Paindaveine, D. (2002a). Multivariate signed ranks: Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1-Norm and Related Methods (Neuchâtel, 2002)*, Stat. Ind. Technol., pages 271–282. Birkhäuser, Basel.

Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli*, 8(6):787–815.

Hallin, M. and Paindaveine, D. (2002c). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, 30(4):1103–1133.

Hallin, M. and Paindaveine, D. (2008). Chernoff-Savage and Hodges-Lehmann results for Wilks' test of multivariate independence. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, volume 1 of *Inst. Math. Stat. (IMS) Collect.*, pages 184–196. Inst. Math. Statist., Beachwood, OH.

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104(4):813–828.

Han, F. and Liu, H. (2018). ECA: high-dimensional elliptical component analysis in non-Gaussian distributions. *J. Amer. Statist. Assoc.*, 113(521):252–268.

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(3):1787–1833.

Hashorva, E., Korshunov, D., and Piterbarg, V. I. (2015). Asymptotic expansion of Gaussian chaos via probabilistic approach. *Extremes*, 18(3):315–347.

Heller, R., Gorfine, M., and Heller, Y. (2012). A class of multivariate distribution-free tests of independence based on graphs. *J. Statist. Plann. Inference*, 142(12):3097–3106.

Heller, R. and Heller, Y. (2016a). Multivariate tests of association based on univariate tests. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 208–216. Curran Associates, Inc.

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

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(29):1–54.

Heller, Y. and Heller, R. (2016b). Computing the Bergsma Dassios sign-covariance. Available at arXiv:1605.08732v1.

Hoeffding, W. (1940). Maszstabinvariante 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(4):546–557.

Hollander, M., Wolfe, D. A., and Chicken, E. (2014). *Nonparametric statistical methods* (3rd ed.). Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

Huang, C. and Huo, X. (2017). A statistically and numerically efficient independence test based on random projections and distance covariance. Available at arXiv:1701.06054v1.

Huo, X. and Székely, G. J. (2016). Fast computing for distance covariance. *Technometrics*, 58(4):435–447.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.*, 14(2):865–880.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, 41(4):2029–2074.

Johnson, O. and Barron, A. (2004). Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, 129(3):391–409.

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

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

Kim, I., Balakrishnan, S., and Wasserman, L. (2020a). Minimax optimality of permutation tests. Available at arXiv:2003.13208v1.

Kim, I., Balakrishnan, S., and Wasserman, L. (2020b). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.*, 48(6):3417–3441.

Kim, I., Balakrishnan, S., and Wasserman, L. (2020c). Supplement to "Robust multivariate nonparametric tests via projection averaging". *Ann. Statist.*, 48(6).

Kinney, J. B. and Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA*, 111(9):3354–3359.

Knight, W. R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *J. Am. Stat. Assoc.*, 61:436–439.

Konijn, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.*, 27(2):300–323.

Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-statistics* (P. V. Malyshev and D. V. Malyshev, Trans.), volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, Netherlands.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Kössler, W. and Rödel, E. (2007). The asymptotic efficacies and relative efficiencies of various linear rank tests for independence. *Metrika*, 65(1):3–28.

Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts* (2nd ed.). Springer Series in Statistics. Springer-Verlag, New York.

Lee, A. J. (1990). *U-Statistics: Theory and Practice*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, NY.

Leeb, H. and Pötscher, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(2):338–376.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). Springer Texts in Statistics. Springer, New York.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.*, 46(1):280–307.

Lin, J. (2017). *Copula versions of RKHS-based and distance-based criteria*. PhD thesis, Pennsylvania State University. Available at https://etda.libraries.psu.edu/catalog/14485jul268.

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88(421):252–260.

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

Lyons, R. (2018). Errata to "Distance covariance in metric spaces". *Ann. Probab.*, 46(4):2400–2405.

Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M., editors (2019). *Handbook of graphical models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL.

Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Adaptive test of independence based on hsic measures. Available at arXiv:1902.06441v4.

Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsbl. Math. Statist.*, 8:234–235.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, 1(4):700–709.

Nandy, P., Weihs, L., and Drton, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Stat.*, 10(2):2287–2311.

Nelsen, R. B. (2006). *An Introduction to Copulas* (2nd ed.). Springer Series in Statistics. Springer, New York, NY.

Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, Cambridge, United Kingdom.

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*, volume 199 of *Lecture Notes in Statistics*. Springer, New York.

Oja, H., Paindaveine, D., and Taskinen, S. (2016). Affine-invariant rank tests for multivariate independence in independent component models. *Electron. J. Stat.*, 10(2):2372–2419.

Orey, S. (1958). A central limit theorem for m-dependent random variables. *Duke Math. J.*, 25(4):543–546.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proc. Roy. Soc. London*, 58:240–242.

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(1):5–31.

Pinelis, I. and Molzon, R. (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electron. J. Stat.*, 10(1):1001–1063.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, Inc., New York-London-Sydney.

Puri, M. L., Sen, P. K., and Gokhale, D. V. (1970). On a class of rank order tests for independence in multivariate distributions. *Sankhyā Ser. A*, 32(3):271–298.

Roy, S. N. (1957). *Some aspects of multivariate analysis*. John Wiley and Sons Inc., New York; Indian Statistical Institute, Calcutta.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.*, 9(4):879–885.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, 41(5):2263–2291.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, NY.

Shah, R. D. and Peters, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.*, 48(3):1514–1538.

Shao, Q.-M. and Zhou, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli*, 22(4):2029–2079.

Sharathkumar, R. and Agarwal, P. K. (2012). Algorithms for the transportation problem in geometric settings. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 306–317, New York, NY. ACM.

Shi, H., Drton, M., and Han, F. (2021+a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* (in press).

Shi, H., Drton, M., and Han, F. (2021+b). On the power of Chatterjee's rank correlation. *Biometrika*. (in press).

Shi, H., Hallin, M., Drton, M., and Han, F. (2020). On universally consistent and fully distribution-free rank tests of vector independence. Available at arXiv:2007.02186v2.

Shorack, G. R. (2017). *Probability for statisticians* (2nd ed.). Springer Texts in Statistics. Springer, Cham, Switzerland.

Spearman, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.*, 15(1):72–101.

Spearman, C. (1906). 'Footrule' for measuring correlation. *Brit. J. Psychol.*, 2(1):89–108.

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.*, 3(4):1236–1265.

Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.*, 117:193–213.

Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.*, 42(6):2382–2412.

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

Taskinen, S., Kankainen, A., and Oja, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.*, 62(1):9–21.

Taskinen, S., Kankainen, A., and Oja, H. (2004). Rank scores tests of multivariate independence. In *Theory and applications of recent robust methods*, Stat. Ind. Technol., pages 329–341. Birkhäuser, Basel.

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

Tomizawa, N. (1971). On some techniques useful for solution of transportation network problems. *Networks*, 1:173–194.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation* (V. Zaiats, Trans.). Springer Series in Statistics. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, United Kingdom.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York.

Vershynin, R. (2018). *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Weihs, L. (2019). *TauStar: Efficient Computation and Testing of the Bergsma-Dassios Sign Covariance*. R package version 1.1.4.

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

Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 105(3):547–562.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3(3):309–326.

Yanagimoto, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.*, 22(1):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(3):455–480.

Zaĭtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Theory Related Fields*, 74(4):535–566.

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.*, 359(11):5345–5363.

Zhu, L., Xu, K., Li, R., and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika*, 104(4):829–843.

Zolotarev, V. M. (1962). Concerning a certain probability problem. *Theory Probab. Appl.*, 6(2):201–204.

Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, 34(6):2879–2896.

Appendix A

APPENDIX OF CHAPTER 2

A.1 Technical proofs

We first introduce more notation. For $x \in \mathbb{R}$, let x_+ denote the positive part of x, defined as $\max\{x, 0\}$. For any vector $\boldsymbol{v} \in \mathbb{R}^p$, we denote $\|\boldsymbol{v}\|$ as its Euclidean norm. We define the L^∞ norm of a random variable as $\|X\|_\infty = \inf\{t \geq 0 : |X| \leq t \text{ a.s.}\}$, the ψ_2 (sub-gaussian) norm as $\|X\|_{\psi_2} = \inf\{t > 0 : \mathrm{E}\exp(X^2/t^2) \leq 2\}$, and the ψ_1 (sub-exponential) norm as $\|X\|_{\psi_1} = \inf\{t > 0 : \mathrm{E}\exp(|X|/t) \leq 2\}$. For any measure P_Z and kernel h, we let $H_n^{(\ell)}(\cdot; \mathrm{P}_Z)$ be the U-statistic based on the completely degenerate kernel $h^{(\ell)}(\cdot; \mathrm{P}_Z)$ from (2.2.2):

$$H_n^{(\ell)}(\cdot; \mathrm{P}_Z) := \binom{n}{\ell}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_\ell \leq n} h^{(\ell)}\Big(Z_{i_1}, \ldots, Z_{i_\ell}; \mathrm{P}_Z\Big). \tag{A.1.1}$$

A.1.1 Proofs for Section 2.3 of the main paper

Proof of Proposition 2.3.2. Since $Y_1, \ldots, Y_d$ are i.i.d. realizations of ζ, we have

$$\mathrm{P}\Big(\max_{j \in [\![d]\!]} Y_j \leq x\Big) = \{\mathrm{P}(\zeta \leq x)\}^d = \{F_\zeta(x)\}^d = \{1 - \overline{F}_\zeta(x)\}^d, \tag{A.1.2}$$

where

$$\overline{F}_\zeta(x) := \mathrm{P}(\zeta > x) = \frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1} \exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1 + o(1)\} \tag{A.1.3}$$

for $x > -\Lambda$ as $x \to \infty$ by Equation (6) in Zolotarev (1962). Take

$$x = 4\lambda_1 \log p + \lambda_1(\mu_1 - 2)\log\log p - \Lambda + \lambda_1 y.$$

Noticing that $x \to \infty$ as $p \to \infty$ and recalling $d = p(p-1)/2$, we obtain

$$\begin{aligned}
&d \cdot \overline{F}_\zeta(x)\\
&= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}\\
&= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}(2\log p)^{\mu_1/2-1}\exp\Big\{-2\log p - \Big(\frac{\mu_1}{2}-1\Big)\log\log p - \frac{y}{2}\Big\}\{1+o(1)\}\\
&= \frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\{1+o(1)\}. \qquad \text{(A.1.4)}
\end{aligned}$$

Combing (A.1.2) and (A.1.4), we deduce that

$$\mathrm{P}(\max_{j\in[\![d]\!]} Y_j \le x) = \{1-\overline{F}_\zeta(x)\}^d \to \exp\Big\{-\lim_{d\to\infty} d\cdot\overline{F}_\zeta(x)\Big\} = \exp\Big\{-\frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\Big\},$$

which concludes the proof of the lemma. □

A.1.2 Proofs for Section 2.4 of the main paper

A.1.2.1 Proof of Theorem 2.4.1

Proof of Theorem 2.4.1. We proceed in two steps, proving first the case $m = 2$ and then generalizing to $m \ge 2$. For notational convenience we introduce the constants $b_1 := \|h\|_\infty < \infty$ and $b_2 := \sup_v \|\phi_v\|_\infty < \infty$.

Step I. Suppose $m = 2$. We start with the scenario that there are infinitely many nonzero eigenvalues. For a large enough integer K to be specified later, we define the "truncated" kernel of $h_2(z_1, z_2; \mathrm{P}_Z)$ as $h_{2,K}(z_1, z_2; \mathrm{P}_Z) = \sum_{v=1}^K \lambda_v \phi_v(z_1)\phi_v(z_2)$, with corresponding U-statistic

$$\widehat{U}_{K,n} := \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_{2,K}(Z_i, Z_j; \mathrm{P}_Z).$$

For simpler presentation, define $Y_{v,i} = \phi_v(Z_i)$ for all $v = 1, 2, \ldots$ and $i \in [\![n]\!]$. In view of the expansions of $h_{2,K}(\cdot)$ and $h_2(\cdot)$, $\widehat{U}_{K,n}$ and $\widehat{U}_n$ can be written as

$$\widehat{U}_{K,n} = \frac{1}{n-1}\Big\{\sum_{v=1}^{K}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2 - \sum_{v=1}^{K}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big)\Big\}$$
$$\text{and}\quad \widehat{U}_n = \frac{1}{n-1}\Big\{\sum_{v=1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2 - \sum_{v=1}^{\infty}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big)\Big\}.$$

We now quantify the approximation accuracy of $\widehat{U}_{K,n}$ to $\widehat{U}_n$. Using Slutsky's argument, we obtain,

$$\begin{aligned}
&\mathrm{P}\Big\{(n-1)\widehat{U}_n \ge x_n\Big\} = \mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2 - \sum_{v=1}^{\infty}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big) \ge x_n\Big\}\\
&\le \mathrm{P}\Big\{\sum_{v=1}^{K}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2 - \sum_{v=1}^{K}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big) \ge x_n - \epsilon_1\Big\}\\
&\quad + \mathrm{P}\Big\{\Big|(n-1)(\widehat{U}_n - \widehat{U}_{K,n})\Big| \ge \epsilon_1\Big\}\\
&\le \mathrm{P}\Big\{\sum_{v=1}^{K}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2 - \sum_{v=1}^{K}\lambda_v \ge x_n - \epsilon_1 - \epsilon_2\Big\} + \mathrm{P}\Big\{\Big|(n-1)(\widehat{U}_n - \widehat{U}_{K,n})\Big| \ge \epsilon_1\Big\}\\
&\quad + \mathrm{P}\Big\{\Big|\sum_{v=1}^{K}\lambda_v\frac{\sum_{i=1}^{n}(Y_{v,i}^2-1)}{n}\Big| \ge \epsilon_2\Big\}, \qquad (\text{A.1.5})
\end{aligned}$$

where ϵ_1, ϵ_2 are constants to be specified later.

The first term on the right-hand side of (A.1.5) may be controlled using Zaĭtsev's multivariate moderate deviation theorem. For this, we require a dimension-free bound on $\sum_{v=1}^{K} u_v\lambda_v^{1/2}(n^{-1/2}Y_{v,i})$ for any $\boldsymbol{u} = (u_1, \ldots, u_K)^\top \in \mathbb{R}^K$ satisfying $\|\boldsymbol{u}\| = 1$. Indeed, we have

$$\Big\|\sum_{v=1}^{K}u_v\lambda_v^{1/2}\frac{Y_{v,i}}{n^{1/2}}\Big\|_\infty \le \sum_{v=1}^{K}|u_v|\lambda_v^{1/2}\frac{\|Y_{v,i}\|_\infty}{n^{1/2}} \le \Big(\sum_{v=1}^{K}u_v^2\Big)^{1/2}\Big(\sum_{v=1}^{K}\lambda_v\Big)^{1/2}n^{-1/2}b_2 \le n^{-1/2}\Lambda^{1/2}b_2.$$

Thus all assumptions in Theorem 1.1 in Zaĭtsev (1987) are satisfied with the τ in his Equation (1.5) chosen to be $n^{-1/2}\Lambda^{1/2}b_2$. We obtain the following bound:

$$\begin{aligned}
&\mathrm{P}\Big\{\sum_{v=1}^{K}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2-\sum_{v=1}^{K}\lambda_v\geq x_n-\epsilon_1-\epsilon_2\Big\}\\
&=\mathrm{P}\Big[\Big\{\sum_{v=1}^{K}\Big(\lambda_v^{1/2}\sum_{i=1}^{n}n^{-1/2}Y_{v,i}\Big)^2\Big\}^{1/2}\geq\Big(x_n-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)_+^{1/2}\Big]\\
&\leq\mathrm{P}\Big[\Big\{\sum_{v=1}^{K}(\lambda_v^{1/2}\xi_v)^2\Big\}^{1/2}\geq\Big(x_n-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)_+^{1/2}-\epsilon_3\Big]\\
&\quad+c_1K^{5/2}\exp\Big\{-\frac{\epsilon_3}{c_2K^{5/2}(n^{-1/2}\Lambda^{1/2}b_2)}\Big\}\\
&=\mathrm{P}\Big[\sum_{v=1}^{K}\lambda_v\xi_v^2\geq\Big\{\Big(x_n-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)_+^{1/2}-\epsilon_3\Big\}_+^2\Big]+c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big),
\end{aligned}\tag{A.1.6}$$

where ϵ_3 is a constant to be specified later. Combining (A.1.5) and (A.1.6), we find using Slutsky's argument once again that

$$\begin{aligned}
&\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2-\sum_{v=1}^{\infty}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big)\geq x_n\Big\}\\
&\leq\mathrm{P}\Big[\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)\geq\Big\{\Big(x_n-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)_+^{1/2}-\epsilon_3\Big\}_+^2-\sum_{v=1}^{K}\lambda_v-\epsilon_4\Big]\\
&\quad+\mathrm{P}\Big\{\Big|(n-1)(\widehat{U}_n-\widehat{U}_{K,n})\Big|\geq\epsilon_1\Big\}+\mathrm{P}\Big\{\Big|\sum_{v=1}^{K}\lambda_v\frac{\sum_{i=1}^{n}(Y_{v,i}^2-1)}{n}\Big|\geq\epsilon_2\Big\}\\
&\quad+c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big)+\mathrm{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2-1)\Big|\geq\epsilon_4\Big\},
\end{aligned}\tag{A.1.7}$$

where ϵ_4 is another constant to be specified later. In the following, we separately study the five terms on the right-hand side of (A.1.7), starting from the first term.

Let $\epsilon_n^* := x_n - [\{(x_n - \epsilon_1 - \epsilon_2 + \sum_{v=1}^K \lambda_v)_+^{1/2} - \epsilon_3\}_+^2 - \sum_{v=1}^K \lambda_v - \epsilon_4]$. Then

$$\epsilon_n^* = \begin{cases} \epsilon_1 + \epsilon_2 + 2\epsilon_3(x_n - \epsilon_1 - \epsilon_2 + \sum_{v=1}^K \lambda_v)^{1/2} - \epsilon_3^2 + \epsilon_4, & \text{if } x_n + \sum_{v=1}^K \lambda_v \geq \epsilon_1 + \epsilon_2 + \epsilon_3^2, \\ x_n + \sum_{v=1}^K \lambda_v + \epsilon_4, & \text{otherwise,} \end{cases}$$

and

$$\mathrm{P}\Big\{\sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1) \geq x_n - \epsilon_n^*\Big\} \leq \mathrm{P}\Big\{\sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1) \geq x_n\Big\} + (\epsilon_n^*)_+ \cdot \max_{x' \in [x_n - (\epsilon_n^*)_+, x_n]} p_\zeta(x') \tag{A.1.8}$$

where $p_\zeta(x)$ is the density of the random variable $\zeta := \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1)$.

We turn to the second term in (A.1.7). Proposition 2.6.1 and Example 2.5.8 in Vershynin (2018) yield that

$$\Big\|n^{-1/2}\sum_{i=1}^n Y_{v,i}\Big\|_{\psi_2}^2 \leq 8n^{-1}\sum_{i=1}^n \Big\|Y_{v,i}\Big\|_{\psi_2}^2 \leq 8(\log 2)^{-1} b_2^2 \leq 12 b_2^2.$$

Applying the triangle inequality and Lemma 2.7.6 in Vershynin (2018), we deduce that

$$\begin{aligned}\Big\|\sum_{v=K+1}^{\infty} \lambda_v\Big(n^{-1/2}\sum_{i=1}^n Y_{v,i}\Big)^2\Big\|_{\psi_1} &\leq \sum_{v=K+1}^{\infty} \lambda_v \Big\|\Big(n^{-1/2}\sum_{i=1}^n Y_{v,i}\Big)^2\Big\|_{\psi_1} \\ &= \sum_{v=K+1}^{\infty} \lambda_v \Big\|n^{-1/2}\sum_{i=1}^n Y_{v,i}\Big\|_{\psi_2}^2 \leq 12 b_2^2 \sum_{v=K+1}^{\infty} \lambda_v.\end{aligned}$$

Using Proposition 2.7.1 in Vershynin (2018), this is seen to further imply that, for any $\epsilon_1' > 0$,

$$\mathrm{P}\Big\{\sum_{v=K+1}^{\infty} \lambda_v\Big(n^{-1/2}\sum_{i=1}^n Y_{v,i}\Big)^2 \geq \epsilon_1'\Big\} \leq 2\exp\Big(-\frac{\epsilon_1'}{12 b_2^2 \sum_{v=K+1}^{\infty} \lambda_v}\Big).$$

Noting that

$$\Big|\sum_{v=K+1}^{\infty} \lambda_v\Big(n^{-1}\sum_{i=1}^n Y_{v,i}^2\Big)\Big| \leq b_2^2 \sum_{v=K+1}^{\infty} \lambda_v,$$

we obtain, for any $\epsilon_1 > b_2^2 \sum_{v=K+1}^{\infty} \lambda_v$,

$$
\begin{aligned}
&\mathrm{P}\Big\{\Big|(n-1)(\widehat{U}_n - \widehat{U}_{K,n})\Big| \geq \epsilon_1\Big\} \\
&\leq \mathrm{P}\Big\{\Big|\sum_{v=K+1}^{\infty} \lambda_v \Big(n^{-1/2}\sum_{i=1}^{n} Y_{v,i}\Big)^2\Big| + \Big|\sum_{v=K+1}^{\infty} \lambda_v \Big(n^{-1}\sum_{i=1}^{n} Y_{v,i}^2\Big)\Big| \geq \epsilon_1\Big\} \\
&\leq \mathrm{P}\Big\{\Big|\sum_{v=K+1}^{\infty} \lambda_v \Big(n^{-1/2}\sum_{i=1}^{n} Y_{v,i}\Big)^2\Big| \geq \epsilon_1 - b_2^2 \sum_{v=K+1}^{\infty} \lambda_v\Big\} \\
&\leq 2e^{1/12} \exp\Big(-\frac{\epsilon_1}{12 b_2^2 \sum_{v=K+1}^{\infty} \lambda_v}\Big). \qquad \text{(A.1.9)}
\end{aligned}
$$

We next study the third term in (A.1.7). Again, Proposition 2.6.1 and Example 2.5.8 in Vershynin (2018) give

$$
\Big\|n^{-1}\sum_{i=1}^{n}(Y_{v,i}^2 - 1)\Big\|_{\psi_2}^2 \leq 8n^{-2}\sum_{i=1}^{n}\Big\|Y_{v,i}^2 - 1\Big\|_{\psi_2}^2 \leq 12 n^{-1}(b_2^2+1)^2,
$$

which further yields

$$
\Big\|\sum_{v=1}^{K}\lambda_v \sum_{i=1}^{n}\frac{Y_{v,i}^2 - 1}{n}\Big\|_{\psi_2} \leq \sum_{v=1}^{K}\lambda_v \Big\|n^{-1}\sum_{i=1}^{n}(Y_{v,i}^2 - 1)\Big\|_{\psi_2} \leq 12^{1/2} n^{-1/2}\Lambda(b_2^2+1).
$$

Using Proposition 2.5.2 in Vershynin (2018), we have, for any $\epsilon_2 > 0$,

$$
\mathrm{P}\Big(\Big|\sum_{v=1}^{K}\lambda_v \sum_{i=1}^{n}\frac{Y_{v,i}^2 - 1}{n}\Big| \geq \epsilon_2\Big) \leq 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\}. \qquad \text{(A.1.10)}
$$

The fourth term in (A.1.7) is explicit, and it remains to bound the fifth and last term. Since ξ_v is a sub-gaussian random variable, $\xi_v^2 - 1$ is sub-exponential. One readily verifies $\|\xi_v^2 - 1\|_{\psi_1} \leq 4$, and accordingly

$$
\Big\|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2 - 1)\Big\|_{\psi_1} \leq \sum_{v=K+1}^{\infty}\lambda_v\|\xi_v^2 - 1\|_{\psi_1} \leq 4\sum_{v=K+1}^{\infty}\lambda_v.
$$

By Proposition 2.7.1 in Vershynin (2018), this further implies that, for any $\epsilon_4 > 0$,

$$\mathrm{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2-1)\Big| \geq \epsilon_4\Big\} \leq 2\exp\Big(-\frac{\epsilon_4}{4\sum_{v=K+1}^{\infty}\lambda_v}\Big). \tag{A.1.11}$$

We now specify the integer K to be $\lfloor n^{(1-3\theta)/5}\rfloor$. By the definition of θ, there exists a positive absolute constant C_θ such that $\sum_{v=K+1}^{\infty}\lambda_v \leq C_\theta n^{-\theta}$ for all sufficiently large n. Combining this fact and inequalities (A.1.7)–(A.1.11), we obtain

$$\begin{aligned}
&\frac{\mathrm{P}\Big\{(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n\Big\}} - 1 \\
&\leq \{\overline{F}_\zeta(x_n)\}^{-1}\Big[(\epsilon_n^*)_+ \cdot \max_{x'\in[x_n-(\epsilon_n^*)_+,x_n]} p_\zeta(x') + 2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12b_2^2\sum_{v=K+1}^{\infty}\lambda_v}\Big) \\
&\quad + 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} + c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big) + 2\exp\Big(-\frac{\epsilon_4}{4\sum_{v=K+1}^{\infty}\lambda_v}\Big)\Big] \\
&\leq \{\overline{F}_\zeta(x_n)\}^{-1}\Big[(\epsilon_n^*)_+ \cdot \max_{x'\in[x_n-(\epsilon_n^*)_+,x_n]} p_\zeta(x') + 2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12b_2^2C_\theta n^{-\theta}}\Big) \\
&\quad + 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} + c_1n^{(1-3\theta)/2}\exp\Big\{-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2n^{(1-3\theta)/2}}\Big\} + 2\exp\Big(-\frac{\epsilon_4}{4C_\theta n^{-\theta}}\Big)\Big],
\end{aligned} \tag{A.1.12}$$

which we shall prove to converge to 0 uniformly on $[-\Lambda, e_n n^\theta]$. The starting point for proving this are Equations (5) and (6) in Zolotarev (1962), which yield that the density $p_\zeta(x)$ and the survival function $\overline{F}_\zeta(x) := \mathrm{P}(\zeta > x)$ of $\zeta = \sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)$ satisfy

$$\begin{aligned}
p_\zeta(x) &= \frac{\kappa}{2\lambda_1\cdot\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\} \\
\text{and}\quad \overline{F}_\zeta(x) &= \frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}
\end{aligned}$$

for $x > -\Lambda$ tending to infinity. Here μ_1 is the multiplicity of the largest eigenvalue λ_1 and $\kappa := \prod_{v=\mu_1+1}^{\infty}(1-\lambda_v/\lambda_1)^{-1/2}$.

Consider the first term in (A.1.12). We claim that there exists an absolute constant

$C_\zeta^* > 0$ such that, for all $0 < \epsilon \le \lambda_1/2$,

$$\sup_{x\ge -\Lambda}\Big|\{\overline{F}_\zeta(x)\}^{-1}\cdot \max_{x'\in[x-\epsilon,x]} p_\zeta(x')\Big| \le C_\zeta^*. \tag{A.1.13}$$

Indeed, we have $p_\zeta(x)/\overline{F}_\zeta(x) = (2\lambda_1)^{-1}\{1+o(1)\}$ as $x\to\infty$, and thus there exists an absolute constant $x_0 > -\Lambda$ such that $p_\zeta(x)/\overline{F}_\zeta(x) \le \lambda_1^{-1}$ for all $x \ge x_0$. Then for all $0<\epsilon\le\lambda_1/2$ and all $x \ge x_0+\epsilon$,

$$\frac{\max_{x'\in[x-\epsilon,x]} p_\zeta(x')}{\overline{F}_\zeta(x)} = \frac{p_\zeta(x-\epsilon')}{\overline{F}_\zeta(x)} \le \frac{p_\zeta(x-\epsilon')}{\overline{F}_\zeta(x-\epsilon')-\epsilon'\cdot p_\zeta(x-\epsilon')} = \frac{1}{\overline{F}_\zeta(x-\epsilon')/p_\zeta(x-\epsilon')-\epsilon'} \le \frac{2}{\lambda_1}$$

where $\epsilon' \in [0,\epsilon]$ is chosen such that $p_\zeta(x-\epsilon') = \max_{x'\in[x-\epsilon,x]} p_\zeta(x')$. Now (A.1.13) holds when taking

$$C_\zeta^* = \max\Big\{\frac{2}{\lambda_1}, \{\overline{F}_\zeta(x_0+\lambda_1/2)\}^{-1}\cdot \max_{x'\in[-\Lambda,x_0+\lambda_1/2]} p_\zeta(x')\Big\}.$$

From (A.1.13), to control the first term in (A.1.12), it remains to show that $(\epsilon_n^*)_+$ converges to 0 uniformly on $[-\Lambda, e_n n^\theta]$ as $n\to\infty$. Choosing

$$\epsilon_1 = 12b_2^2C_\theta n^{-\theta}\Big(\frac{x_n+\Lambda}{2\lambda_1}+n^{\theta/2}\Big),\quad \epsilon_2 = n^{-\theta},\quad \epsilon_3 = n^{-\theta/2},\quad \epsilon_4 = 4C_\theta n^{-\theta}\Big(\frac{x_n+\Lambda}{2\lambda_1}+n^{\theta/2}\Big), \tag{A.1.14}$$

we deduce that the first term in (A.1.12) converges uniformly to 0 on $[-\Lambda, e_n n^\theta]$ as $n\to\infty$ by observing that if $x_n + \sum_{v=1}^K \lambda_v \ge \epsilon_1+\epsilon_2+\epsilon_3^2$,

$$\begin{aligned}\epsilon_n^* &\le \epsilon_1+\epsilon_2+2\epsilon_3(x_n+\Lambda)^{1/2}-\epsilon_3^2+\epsilon_4\\ &\le \frac{6b_2^2C_\theta+2C_\theta}{\lambda_1}\Big(e_n+\frac{\Lambda}{n^\theta}\Big)+2\Big(e_n+\frac{\Lambda}{n^\theta}\Big)^{1/2}+(12b_2^2C_\theta+4C_\theta)n^{-\theta/2},\end{aligned}$$

and otherwise

$$\epsilon_n^* \le \epsilon_1+\epsilon_2+\epsilon_3^2+\epsilon_4 \le \frac{6b_2^2C_\theta+2C_\theta}{\lambda_1}\Big(e_n+\frac{\Lambda}{n^\theta}\Big)+(12b_2^2C_\theta+4C_\theta)n^{-\theta/2}+2n^{-\theta}.$$

Recall that we consider a positive sequence $\{e_n\}$ tending to 0.

We then further verify that the other four terms in (A.1.12) also converge to 0 uniformly on $[-\Lambda, e_n n^\theta]$ as $n \to \infty$. There exists some absolute constant $c_\zeta^* > 0$ such that for all $x \geq 2\lambda_1 - \Lambda$,

$$\overline{F}_\zeta(x) \geq c_\zeta^* \frac{\kappa}{\Gamma(\mu_1/2)} \Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1} \exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big). \tag{A.1.15}$$

We then have, by noticing $\theta < 1/3$, for all n large enough and all $x_n \in [2\lambda_1 - \Lambda, e_n n^\theta]$,

$$\begin{aligned}
&\{\overline{F}_\zeta(x_n)\}^{-1} \exp\Big(-\frac{\epsilon_1}{12 b_2^2 C_\theta n^{-\theta}}\Big) \leq \frac{\Gamma(\mu_1/2)}{c_\zeta^* \kappa}\Big(\frac{e_n n^\theta + \Lambda}{2\lambda_1}\Big)^{1/2} \exp(-n^{\theta/2}),\\
&\{\overline{F}_\zeta(x_n)\}^{-1} \exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} \leq \frac{\Gamma(\mu_1/2)}{c_\zeta^* \kappa}\Big(\frac{e_n n^\theta + \Lambda}{2\lambda_1}\Big)^{1/2} \exp(-C' n^{1/3}),\\
&\{\overline{F}_\zeta(x_n)\}^{-1} n^{(1-3\theta)/2} \exp\Big\{-\frac{n^{1/2}\epsilon_3}{c_2 \Lambda^{1/2} b_2 n^{(1-3\theta)/2}}\Big\} \leq \frac{\Gamma(\mu_1/2)}{c_\zeta^* \kappa}\Big(\frac{e_n n^\theta + \Lambda}{2\lambda_1}\Big)^{1/2} n^{(1-3\theta)/2} \exp(-C'' n^\theta),\\
&\{\overline{F}_\zeta(x_n)\}^{-1} \exp\Big(-\frac{\epsilon_4}{4 C_\theta n^{-\theta}}\Big) \leq \frac{\Gamma(\mu_1/2)}{c_\zeta^* \kappa}\Big(\frac{e_n n^\theta + \Lambda}{2\lambda_1}\Big)^{1/2} \exp(-n^{\theta/2}).
\end{aligned} \tag{A.1.16}$$

Here C' and C'' are some absolute positive constants. The inequalities in (A.1.16) hold for all sufficiently large n and all $x_n \in [-\Lambda, 2\lambda_1 - \Lambda]$ with replacing $\Gamma(\mu_1/2)/c_\zeta^*\kappa \cdot \{(e_n n^\theta + \Lambda)/(2\lambda_1)\}^{1/2}$ by $\{\overline{F}_\zeta(2\lambda_1 - \Lambda)\}^{-1}$, which together with (A.1.16) concludes the uniform convergence.

If there are only finitely many nonzero eigenvalues, a simple modification to (A.1.12) gives

$$\begin{aligned}
\frac{\mathrm{P}\Big\{(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x_n\Big\}} - 1 \leq{}& \frac{1}{\overline{F}_\zeta(x)}\Big[(\epsilon_n^*)_+ \cdot \max_{x' \in [x_n - (\epsilon_n^*)_+, x_n]} p_\zeta(x')\\
&+ 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} + c_1 K^{5/2} \exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2} b_2 K^{5/2}}\Big)\Big],
\end{aligned} \tag{A.1.17}$$

where $\epsilon_n^* := x_n - [\{(x_n - \epsilon_2 + \Lambda)_+^{1/2} - \epsilon_3\}_+^2 - \Lambda]$ and K is the number of nonzero eigenvalues. Choosing $\epsilon_2 = n^{-1/3}$, $\epsilon_3 = n^{-1/6}$, one can obtain that the right-hand side of (A.1.17) converges uniformly to 0 on $[-\Lambda, e_n n^\theta]$ as $n \to \infty$.

We thus proved that

$$\sup_{x_n \in [-\Lambda, e_n n^\theta]} \left[\frac{\mathrm{P}\Big\{(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x_n\Big\}} - 1 \right] \le o(1).$$

For the lower bound, it can be shown similarly that if there are infinitely many nonzero eigenvalues, then

$$\begin{aligned} &\frac{\mathrm{P}\Big\{(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x_n\Big\}} - 1 \\ &\ge \{\overline{F}_\zeta(x_n)\}^{-1}\Big[-(\epsilon_n^{**})_+ \cdot \max_{x'' \in [x_n, x_n + (\epsilon_n^{**})_+]} p_\zeta(x'') - 2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12 b_2^2 C_\theta n^{-\theta}}\Big) \\ &\quad -2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} - c_1 n^{(1-3\theta)/2}\exp\Big\{-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2 n^{(1-3\theta)/2}}\Big\} - 2\exp\Big(-\frac{\epsilon_4}{4C_\theta n^{-\theta}}\Big)\Big], \end{aligned} \tag{A.1.18}$$

where

$$\begin{aligned} \epsilon_n^{**} &:= \Big[\Big\{\Big(x_n + \epsilon_1 + \epsilon_2 + \sum_{v=1}^K \lambda_v\Big)_+^{1/2} + \epsilon_3\Big\}^2 - \sum_{v=1}^K \lambda_v + \epsilon_4\Big] - x_n \\ &= \begin{cases} \epsilon_1 + \epsilon_2 + 2\epsilon_3(x_n + \epsilon_1 + \epsilon_2 + \sum_{v=1}^K \lambda_v)^{1/2} + \epsilon_3^2 + \epsilon_4, & \text{if } x_n + \sum_{v=1}^K \lambda_v \ge -\epsilon_1 - \epsilon_2, \\ -x_n - \sum_{v=1}^K \lambda_v + \epsilon_3^2 + \epsilon_4, & \text{otherwise.} \end{cases} \end{aligned}$$

We choose (A.1.14) as well. To conclude the lower bound, it suffices to notice that there exists an absolute constant $C_\zeta^{**} > 0$ such that

$$\sup_{x \ge -\Lambda} \Big| \{\overline{F}_\zeta(x)\}^{-1} \cdot \max_{x'' \in [x, x + (\epsilon_n^{**})_+]} p_\zeta(x'') \Big| \le C_\zeta^{**},$$

and ϵ_n^{**} converges uniformly to 0 on $[-\Lambda, e_n n^\theta]$: if $x_n + \sum_{v=1}^K \lambda_v \ge -\epsilon_1 - \epsilon_2$, then

$$0 < \epsilon_n^{**} \le \frac{6b_2^2 C_\theta + 2C_\theta}{\lambda_1}\Big(e_n + \frac{\Lambda}{n^\theta}\Big) + 2\Big(e_n + \frac{\Lambda + 2}{n^\theta}\Big)^{1/2} + (12 b_2^2 C_\theta + 4C_\theta) n^{-\theta/2} + 2n^{-\theta}$$

for all n large enough, and otherwise

$$0 < \epsilon_1 + \epsilon_2 + \epsilon_3^2 + \epsilon_4 \le \epsilon_n^{**} \le \sum\nolimits_{v > \lfloor n^{(1-3\theta)/5} \rfloor} \lambda_v + \frac{2C_\theta}{\lambda_1}\Big(e_n + \frac{\Lambda}{n^\theta}\Big) + 4C_\theta n^{-\theta/2} + n^{-\theta}.$$

If there are only finitely many nonzero eigenvalues, one can obtain

$$\frac{\mathrm{P}\Big\{(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x_n\Big\}} - 1 \ge \frac{1}{\overline{F}_\zeta(x)}\Big[-(\epsilon_n^{**})_+ \cdot \max_{x'' \in [x_n, x_n + (\epsilon_n^{**})_+]} p_\zeta(x'') - 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} - c_1 K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big)\Big], \tag{A.1.19}$$

where $\epsilon_n^{**} := [\{(x_n + \epsilon_2 + \Lambda)_+^{1/2} + \epsilon_3\}^2 - \Lambda] - x_n$ and K is the number of nonzero eigenvalues. Choosing $\epsilon_2 = n^{-1/3}$, $\epsilon_3 = n^{-1/6}$, one can verify that the right-hand side of (A.1.19) converges uniformly to 0 on $[-\Lambda, e_n n^\theta]$ as $n \to \infty$. This completes the proof of the case $m = 2$.

Step II. We use the Hoeffding decomposition and the exponential inequality for bounded completely degenerate U-statistics of Arcones and Giné (1993) to prove the general case $m \ge 2$. Write

$$\binom{m}{2}^{-1}(n-1)\widehat{U}_n = (n-1)H_n^{(2)}(\cdot;\mathrm{P}_Z) + \sum_{\ell=3}^m \binom{m}{2}^{-1}\binom{m}{\ell}(n-1)H_n^{(\ell)}(\cdot;\mathrm{P}_Z).$$

Using Slutsky's argument, we have

$$\frac{\mathrm{P}\Big\{\binom{m}{2}^{-1}(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x_n\Big\}} \le \frac{\mathrm{P}\Big\{(n-1)H_n^{(2)}(\cdot;\mathrm{P}_Z) > x_n - \epsilon_n^{\#}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x_n\Big\}} + \sum_{\ell=3}^m \frac{\mathrm{P}\Big\{\binom{m}{2}^{-1}\binom{m}{\ell}(n-1)\cdot|H_n^{(\ell)}(\cdot;\mathrm{P}_Z)| \ge \epsilon_{n,\ell}^{\#}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x_n\Big\}}, \tag{A.1.20}$$

where $\{\epsilon_{n,\ell}^{\#}, \ell = 3, \ldots, m\}$ are constants to be specified later and $\epsilon_n^{\#} := \sum_{\ell=3}^m \epsilon_{n,\ell}^{\#}$.

We analyze the first term and the remaining terms on the right-hand side of (A.1.20) separately. To bound the latter, we employ Proposition 2.3(c) in Arcones and Giné (1993), which states that there exist absolute positive constants C'_ℓ and C''_ℓ such that for all $\epsilon_5 > 0$,

$$\mathrm{P}(n^{\ell/2}|H_n^{(\ell)}(\cdot;\mathrm{P}_Z)| \geq \epsilon_5) \leq C'_\ell \exp\{-C''_\ell(\epsilon_5/\|h^{(\ell)}(\cdot;\mathrm{P}_Z)\|_\infty)^{2/\ell}\}, \tag{A.1.21}$$

where $\|h^{(\ell)}(\cdot;\mathrm{P}_Z)\|_\infty \leq 2^\ell b_1$ can be shown by the alternative formula of $h^{(\ell)}(z_1,\ldots,z_\ell;\mathrm{P}_Z)$ as below:

$$h^{(\ell)}(z_1,\ldots,z_\ell;\mathrm{P}_Z) = h_\ell(z_1,\ldots,z_\ell;\mathrm{P}_Z)+\sum_{k=1}^{\ell-1}(-1)^{\ell-k}\sum_{1\leq i_1<\cdots<i_k\leq\ell} h_k(z_{i_1},\ldots,z_{i_k};\mathrm{P}_Z)+(-1)^\ell\mathrm{E}h.$$

Plugging (A.1.21) into each term in the sum on the right of (A.1.20) implies, for $n \geq 2$,

$$\begin{aligned}
&\sum_{\ell=3}^m \frac{\mathrm{P}\Big\{\binom{m}{2}^{-1}\binom{m}{\ell}(n-1)\cdot|H_n^{(\ell)}(\cdot;\mathrm{P}_Z)| \geq \epsilon^{\#}_{n,\ell}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x_n\Big\}} \\
&\leq \sum_{\ell=3}^m \{\overline{F}_\zeta(x_n)\}^{-1}C'_\ell \exp\Big[-C''_\ell\Big\{n^{\ell/2-1}\binom{m}{2}\binom{m}{\ell}^{-1}\epsilon^{\#}_{n,\ell}\Big/\|h^{(\ell)}(\cdot;\mathrm{P}_Z)\|_\infty\Big\}^{2/\ell}\Big] \\
&\leq \begin{cases}
\displaystyle\sum_{\ell=3}^m \frac{C'_\ell}{c^*_\zeta}\Big\{\Big(\frac{x_n+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x_n+\Lambda}{2\lambda_1}\Big)\Big\}^{-1}\exp\Big[-C''_\ell n^\theta\Big\{\binom{m}{2}\binom{m}{\ell}^{-1}\epsilon^{\#}_{n,\ell}\Big/(2^\ell b_1)\Big\}^{2/\ell}\Big], \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{for } x_n \in [2\lambda_1-\Lambda, e_n n^\theta], \\
\displaystyle\sum_{\ell=3}^m C'_\ell\{\overline{F}_\zeta(2\lambda_1-\Lambda)\}^{-1}\exp\Big[-C''_\ell n^\theta\Big\{\binom{m}{2}\binom{m}{\ell}^{-1}\epsilon^{\#}_{n,\ell}\Big/(2^\ell b_1)\Big\}^{2/\ell}\Big], \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{for } x_n \in [-\Lambda, 2\lambda_1-\Lambda],
\end{cases}
\end{aligned} \tag{A.1.22}$$

where the last step is due to (A.1.15) and the fact that $\theta \leq 1/3 \leq 1-2/\ell$ for $\ell \geq 3$. Taking

$$\epsilon^{\#}_{n,\ell} = b_1\binom{m}{2}^{-1}\binom{m}{\ell}\Big\{\frac{4}{C''_\ell n^\theta}\Big(\frac{x_n+\Lambda}{2\lambda_1}+n^{\theta/2}\Big)\Big\}^{\ell/2},$$

the sum on the right-hand side of (A.1.22) is seen to be $o(1)$. It remains to control the first term in (A.1.20). We start by writing the term as

$$\begin{aligned}
&\frac{\mathrm{P}\Big\{(n-1)H_n^{(2)} > x_n - \epsilon_n^{\#}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n\Big\}} \\
&= \frac{\mathrm{P}\Big\{(n-1)H_n^{(2)} > x_n - \epsilon_n^{\#}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n - \epsilon_n^{\#}\Big\}} \cdot \frac{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n - \epsilon_n^{\#}\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n\Big\}}. \qquad \text{(A.1.23)}
\end{aligned}$$

The first factor in (A.1.23) converges uniformly to 1 on $[-\Lambda, e_n n^{\theta}]$ by going through the same proof in Step I while noticing that although $x_n - \epsilon_n^{\#}$ is not necessarily greater than or equal to $-\Lambda$, it holds for all $x_n \in [-\Lambda, e_n n^{\theta}]$ that

$$0 < \epsilon_n^{\#} = \sum_{\ell=3}^{m} \epsilon_{n,\ell}^{\#} \le \sum_{\ell=3}^{m} b_1 \binom{m}{2}^{-1} \binom{m}{\ell} \Big\{ \frac{2}{C''\lambda_1}\Big(e_n + \frac{\Lambda}{n^{\theta}}\Big) + \frac{4}{C''} n^{-\theta/2} \Big\}^{\ell/2}. \qquad \text{(A.1.24)}$$

For the second term in (A.1.23), we have

$$1 \le \frac{\overline{F}_{\zeta}(x_n - \epsilon_n^{\#})}{\overline{F}_{\zeta}(x_n)} \le 1 + \frac{\epsilon_n^{\#} \cdot \max_{x' \in [x_n - \epsilon_n^{\#}, x_n]} p_{\zeta}(x')}{\overline{F}_{\zeta}(x_n)} \le 1 + C_{\zeta}^{*} \cdot \epsilon_n^{\#}$$

for $x_n > 0$ and $\epsilon_n^{\#} \le \lambda_1/2$ by (A.1.13). By (A.1.24) again, we have the second term in (A.1.23) uniformly converges to 1 as well. Therefore, we obtain the right-hand side of (A.1.20) is uniformly converges to 1 on $[-\Lambda, e_n n^{\theta}]$ as $n \to \infty$. Consequently,

$$\sup_{x_n \in [-\Lambda, e_n n^{\theta}]} \left[\frac{\mathrm{P}\Big\{\binom{m}{2}^{-1}(n-1)\widehat{U}_n > x_n\Big\}}{\mathrm{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1) > x_n\Big\}} - 1 \right] \le o(1).$$

Again a similar derivation yields a corresponding lower bound of order $o(1)$, completing the proof of the general case $m \ge 2$. □

A.1.2.2 Proof of Theorem 2.4.2

Proof of Theorem 2.4.2. Since the marginal distributions are assumed continuous, we may assume, without loss of generality, that they are uniform distributions on $[0,1]$. Theorem 2.4.1 can then directly apply to the studied kernel $h(\cdot)$ in view of Assumption 2.2.1.

The main tool in this proof is Theorem 1 in Arratia et al. (1989). Specifically, we use the version presented in Lemma C2 in Han et al. (2017). We let $I := \{(j,k) : 1 \le j < k \le p\}$, and for all $u := (j,k) \in I$, we define $B_u = \{(\ell, v) \in I : \{\ell, v\} \cap \{j,k\} \neq \varnothing\}$ and

$$\eta_u := \eta_{jk} := \binom{m}{2}^{-1} (n-1)\widehat{U}_{jk}.$$

Then the theorem yields that

$$\left| \mathrm{P}\Big(\max_{u \in I} \eta_u \le t \Big) - \exp(-L_n) \right| \le A_1 + A_2 + A_3, \tag{A.1.25}$$

where $L_n = \sum_{u \in I} \mathrm{P}(\eta_u > t)$,

$$A_1 = \sum_{u \in I} \sum_{\beta \in B_u} \mathrm{P}(\eta_u > t)\mathrm{P}(\eta_\beta > t), \qquad A_2 = \sum_{u \in I} \sum_{\beta \in B_u \setminus \{u\}} \mathrm{P}(\eta_u > t, \eta_\beta > t),$$

$$\text{and} \quad A_3 = \sum_{u \in I} \mathrm{E}|\mathrm{P}\{\eta_u > t \mid \sigma(\eta_\beta : \beta \notin B_u)\} - \mathrm{P}(\eta_u > t)|.$$

We now choose an appropriate value of t such that L_n tends to a constant independent of p as $n \to \infty$. Let

$$t = 4\lambda_1 \log p + \lambda_1(\mu_1 - 2)\log\log p - \Lambda + \lambda_1 y \ \asymp\ 4\lambda_1 \log p = o(n^\theta). \tag{A.1.26}$$

By Theorem 2.4.1,

$$L_n = \frac{p(p-1)}{2}\mathrm{P}(\eta_{12} > t) = \frac{p(p-1)}{2}\overline{F}_\zeta(t)\{1 + o(1)\}. \tag{A.1.27}$$

Using Example 5 in Hashorva et al. (2015), we have for any $t > -\Lambda$,

$$\overline{F}_{\zeta}(t) = \frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{t+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{t+\Lambda}{2\lambda_1}\Big)[1+O\{(\log p)^{-1}\}]. \tag{A.1.28}$$

Combining (A.1.27) and (A.1.28) implies

$$\begin{aligned} L_n &= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{t+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{t+\Lambda}{2\lambda_1}\Big)\{1+o(1)\} \\ &= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}(2\log p)^{\mu_1/2-1}\exp\Big\{-2\log p-\Big(\frac{\mu_1}{2}-1\Big)\log\log p-\frac{y}{2}\Big\}\{1+o(1)\} \\ &= \frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\{1+o(1)\}, \end{aligned} \tag{A.1.29}$$

where $\kappa := \prod_{v=\mu_1+1}(1-\lambda_v/\lambda_1)^{-1/2}$.

Next we bound A_1, A_2, and A_3 separately. We have

$$A_1 = \tfrac{1}{2}p(p-1)(2p-3)\{\mathrm{P}(\eta_{12} > t)\}^2.$$

Moreover, since Hoeffding's D is a rank-based U-statistic, Proposition 2.2.1(ii) yields that η_u is independent of η_β for all $u \in I, \beta \in B_u\backslash\{u\}$. Hence,

$$A_2 = \sum_{u\in I}\sum_{\beta\in B_u\backslash\{u\}}\mathrm{P}(\eta_u > t)\mathrm{P}(\eta_\beta > t) = p(p-1)(p-2)\{\mathrm{P}(\eta_{12} > t)\}^2.$$

Again, by Proposition 2.2.1(iii), we have $A_3 = 0$. Accordingly,

$$A_1 + A_2 + A_3 \le 2p(p-1)^2\{\mathrm{P}(\eta_{12} > t)\}^2 = \frac{2(2L_n)^2}{p} = O\Big(\frac{1}{p}\Big). \tag{A.1.30}$$

Let $L = 2^{\mu_1/2-2}\kappa/\Gamma(\mu_1/2)\cdot\exp(-y/2)$. Plugging (A.1.26), (A.1.29), (A.1.30) into (A.1.25) yields

$$\begin{aligned} &\Big|\mathrm{P}\Big\{\binom{m}{2}^{-1}(n-1)\max_{j<k}\widehat{U}_{jk} - 4\lambda_1\log p - \lambda_1(\mu_1-2)\log\log p + \Lambda \le \lambda_1 y\Big\} - \exp(-L)\Big| \\ &\le \Big|\mathrm{P}\Big(\max_{u\in I}\eta_\alpha \le t\Big) - \exp(-L_n)\Big| + \Big|\exp(-L_n) - \exp(-L)\Big| = o(1). \end{aligned}$$

This completes the proof. □

A.1.2.3 Proof of Corollary 2.4.1

Proof of Corollary 2.4.1. We only give the proof for Hoeffding's D here. The proofs for the other two tests are very similar and hence omitted. As in the proof of Theorem 2.4.2, we may assume the margins to be uniformly distributed on $[0, 1]$ without loss of generality. To employ Theorem 2.4.2, we only need to compute θ. We claim that

$$\sum_{v=K+1}^{\infty} \lambda_v \asymp \frac{(\log K)^2}{K}. \tag{A.1.31}$$

If this claim is true, then by the definition of θ, one obtains $\theta = 1/8 - \delta$, where δ is an arbitrarily small pre-specified positive absolute constant.

We now prove (A.1.31). Notice that the K largest eigenvalues are corresponding to the K smallest products ij, $i, j \in \mathbb{Z}^+$. We begin by assuming that there exists an integer M such that the number of pairs (i, j) satisfying $ij \le M$ is exactly K:

$$2 \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \lfloor M/i \rfloor - \lfloor M^{1/2} \rfloor^2 = K. \tag{A.1.32}$$

To analyze $\sum_{v=K+1}^{\infty} \lambda_v$, we first quantify M. An upper bound on $\sum_{i=1}^{\lfloor M^{1/2} \rfloor} \lfloor M/i \rfloor$ is

$$\sum_{i=1}^{\lfloor M^{1/2} \rfloor} \Big\lfloor \frac{M}{i} \Big\rfloor \le \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \frac{M}{i} = M \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \frac{1}{i} \le M\Big(\log \lfloor M^{1/2} \rfloor + 1 \Big) \le M\Big(\frac{1}{2} \log M + 1 \Big),$$

and a lower bound is

$$\begin{aligned}\sum_{i=1}^{\lfloor M^{1/2} \rfloor} \Big\lfloor \frac{M}{i} \Big\rfloor &\ge \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \Big(\frac{M}{i} - 1 \Big) = M \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \frac{1}{i} - \lfloor M^{1/2} \rfloor \\ &\ge M \log \lfloor M^{1/2} \rfloor - \lfloor M^{1/2} \rfloor \ge M \log(M^{1/2} - 1) - M^{1/2}.\end{aligned}$$

Thus we have $M \log M \asymp K$, which implies that $M \asymp K/\log K$. Then we obtain

$$\begin{aligned}\sum_{v=K+1}^{\infty} \lambda_v &\asymp \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \sum_{j=\lfloor M/i \rfloor+1}^{\infty} \frac{1}{i^2 j^2} + \sum_{j=1}^{\lfloor M^{1/2} \rfloor} \sum_{i=\lfloor M/j \rfloor+1}^{\infty} \frac{1}{i^2 j^2} + \sum_{i=\lfloor M^{1/2} \rfloor+1}^{\infty} \sum_{j=\lfloor M^{1/2} \rfloor+1}^{\infty} \frac{1}{i^2 j^2} \\ &\asymp \sum_{i=1}^{\lfloor M^{1/2} \rfloor} \frac{1}{i^2 (M/i)} + \sum_{j=1}^{\lfloor M^{1/2} \rfloor} \frac{1}{(M/j) j^2} + \frac{1}{(M^{1/2})(M^{1/2})} \\ &\asymp 2\Big\{\frac{\log(M^{1/2})}{M}\Big\} + \frac{1}{M} \asymp \frac{(\log K)^2}{K}.\end{aligned}$$

If there is no integer M such that (A.1.32) holds, then we pick the largest integer M_1 and the smallest integer M_2 such that

$$2 \sum_{i=1}^{\lfloor M_1^{1/2} \rfloor} \Big\lfloor \frac{M_1}{i} \Big\rfloor - \lfloor M_1^{1/2} \rfloor^2 < K < 2 \sum_{i=1}^{\lfloor M_2^{1/2} \rfloor} \Big\lfloor \frac{M_2}{i} \Big\rfloor - \lfloor M_2^{1/2} \rfloor^2,$$

and let K_1 and K_2 denote the left-hand side and the right-hand side, respectively. One can verify that $K_1 > K/2$ and $K_2 < 2K$ for all sufficiently large K. Then we have

$$\begin{aligned} &\sum_{v=K+1}^{\infty} \lambda_v \le \sum_{v=K_1+1}^{\infty} \lambda_v \asymp \frac{(\log K_1)^2}{K_1} \le \frac{2(\log K)^2}{K} \\ \text{and} \quad &\sum_{v=K+1}^{\infty} \lambda_v \ge \sum_{v=K_2+1}^{\infty} \lambda_v \asymp \frac{(\log K_2)^2}{K_2} \ge \frac{(\log K)^2}{2K}.\end{aligned}$$

Therefore, the asymptotic result for $\sum_{v=K+1}^{\infty} \lambda_v$ given in (A.1.31) still holds. □

A.1.2.4 Proof of Lemma 2.4.1

Proof of Lemma 2.4.1. Again we only prove the claim for Hoeffding's D; Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^* can be treated similarly. We first establish the fact that $D_{jk} \asymp \Sigma_{jk}^2$ as $\Sigma_{jk} \to 0$. Let $\{(X_{ji}, X_{ki})^\top : i \in [\![5]\!]\}$ be a collection of independent and identically distributed random vectors that follow a bivariate normal distribution with mean $(0,0)^\top$ and covariance matrix

$$\begin{bmatrix} 1 & \Sigma_{jk} \\ \Sigma_{jk} & 1 \end{bmatrix}.$$

We have

$$D_{jk} = \mathrm{E}_{jk} h_D = \int h_D(x_{j1}, x_{k1}, \ldots, x_{j5}, x_{k5}) \phi(x_{j1}, x_{k1}, \ldots, x_{j5}, x_{k5}; \Sigma_{jk}) \prod_{i=1}^{5} \mathrm{d}x_{ji} \prod_{i=1}^{5} \mathrm{d}x_{ki},$$

where

$$\phi(x_{j1}, x_{k1}, \ldots, x_{j5}, x_{k5}; \Sigma_{jk}) = \prod_{i=1}^{5} \phi(x_{ji}, x_{ki}; \Sigma_{jk}),$$

and

$$\phi(x_{ji}, x_{ki}; \Sigma_{jk}) = \frac{1}{2\pi(1-\Sigma_{jk}^2)^{1/2}} \exp\Big\{ -\frac{x_{ji}^2 + x_{ki}^2 - 2\Sigma_{jk} x_{ji} x_{ki}}{2(1-\Sigma_{jk}^2)} \Big\}$$

is the joint density of $(X_{ji}, X_{ki})^\top$. Notice that D_{jk} is smooth with respect to Σ_{jk}:

$$\frac{\partial^s D_{jk}}{\partial \Sigma_{jk}^s} = \int h_D(x_{j1}, x_{k1}, \ldots, x_{j5}, x_{k5}) \frac{\partial^s \phi(x_{j1}, x_{k1}, \ldots, x_{j5}, x_{k5}; \Sigma_{jk})}{\partial \Sigma_{jk}^s} \prod_{i=1}^{5} \mathrm{d}x_{ji} \prod_{i=1}^{5} \mathrm{d}x_{ki}.$$

In order to prove $D_{jk} \asymp \Sigma_{jk}^2$, it suffices to establish that $D_{jk} = 0$ when $\Sigma_{jk} = 0$, the first derivative of D_{jk} with respect to Σ_{jk} is 0 at $\Sigma_{jk} = 0$, and the second derivative of D_{jk} with respect to Σ_{jk} is $5/\pi^2$ at $\Sigma_{jk} = 0$, which can be confirmed by a lengthy but straightforward computation.

Now we turn to our claim. Recall that $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathrm{P}_{jk})\} = 0$ when $\Sigma_{jk} = 0$. We will show that the first-order term in the Taylor series of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathrm{P}_{jk})\}$ with respect to Σ_{jk} is also 0. Suppose, for contradiction, the first-order coefficient (denoted by a_1) in the Taylor series of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathrm{P}_{jk})\}$ is not 0, then for Σ_{jk} in a sufficiently small neighborhood of 0, $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathrm{P}_{jk})\} < 0$ for $\Sigma_{jk} < 0$ if $a_1 > 0$, and for $\Sigma_{jk} > 0$ if $a_1 < 0$, which contradicts the definition of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathrm{P}_{jk})\}$. This together with $\mathrm{E}_{jk} h_D \asymp \Sigma_{jk}^2$ completes the proof. $\square$

A.1.2.5 Proof of Theorem 2.4.3

Proof of Theorem 2.4.3. It is clear that we only have to consider $\max_{j<k} U_{jk} = C_\gamma(\log p/n)$ for some sufficiently large C_γ. The main idea here is to bound $\max_{j<k} \widehat{U}_{jk} - \max_{j<k} U_{jk}$ with high probability. To do this, we first construct a concentration inequality for $|\widehat{U}_{jk} - U_{jk}|$. The Hoeffding decomposition of the difference is

$$\widehat{U}_{jk} - U_{jk} = \frac{m}{n}\sum_{i=1}^{n} h^{(1)}\{(X_{ji}, X_{ki})^\top; \mathrm{P}_{jk}\} + \sum_{\ell=2}^{m}\binom{m}{\ell} H_n^{(\ell)}(\cdot; \mathrm{P}_{jk}). \tag{A.1.33}$$

For controlling the first term in (A.1.33), recall that $\|h\|_\infty \le b_1 < \infty$, and then $h^{(1)}(\cdot; \mathrm{P}_{jk}) = h_1(\cdot; \mathrm{P}_{jk}) - \mathrm{E}h$ is bounded by $2b_1$ almost surely and $\mathrm{E}h^{(1)}(\cdot; \mathrm{P}_{jk}) = 0$. We then apply Bernstein's inequality, giving

$$\mathrm{P}\Big\{\frac{m}{n}\Big|\sum_{i=1}^{n} h^{(1)}(\cdot; \mathrm{P}_{jk})\Big| > t_1\Big\} \le 2\exp\Big(-\frac{n(t_1/m)^2}{2[\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathrm{P}_{jk})\} + 2b_1(t_1/m)/3]}\Big). \tag{A.1.34}$$

By the definition of the distribution family $\mathcal{D}(\gamma, p; h)$, we have

$$\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathrm{P}_{jk})\} \le \gamma \mathrm{E}_{jk} h = \gamma U_{jk} \le \gamma C_\gamma(\log p/n).$$

Plugging this into (A.1.34) and taking $t_1 = C_1(\log p/n)$, where C_1 is a constant to be specified later, yields

$$\begin{aligned}
&\mathrm{P}\Big\{\frac{m}{n}\Big|\sum_{i=1}^{n} h^{(1)}(\cdot; \mathrm{P}_{jk})\Big| > C_1\frac{\log p}{n}\Big\} \\
&\le 2\exp\Big\{-\frac{C_1^2 \log p}{2(m^2\gamma C_\gamma + 2mb_1C_1/3)}\Big\} = 2\Big(\frac{1}{p}\Big)^{C_1^2/(2m^2\gamma C_\gamma + 4mb_1C_1/3)}.
\end{aligned} \tag{A.1.35}$$

We then handle the remaining term. By Proposition 2.3(c) in Arcones and Giné (1993), there exist absolute constants $C'_\ell, C''_\ell > 0$ such that for all $t \in (0, 1]$, $2 \le \ell \le m$,

$$\mathrm{P}(|H_n^{(\ell)}(\cdot; \mathrm{P}_{jk})| \ge t) \le C'_\ell \exp\Big\{-C''_\ell n\Big(\frac{t}{\|h^{(\ell)}(\cdot; \mathrm{P}_{jk})\|_\infty}\Big)^{2/\ell}\Big\}$$

$$\le C'_\ell \exp\Big\{ - C''_\ell n \Big(\frac{t}{2^\ell b_1}\Big)^{2/\ell}\Big\} \le C'_\ell \exp\Big(- \frac{C''_\ell n t}{4 b_1^{2/\ell}}\Big),$$

which further implies that

$$\mathrm{P}\Big\{\Big|\sum_{\ell=2}^{k}\binom{m}{\ell} H_n^{(\ell)}(\cdot;\mathrm{P}_{jk})\Big| \ge t_2\Big\} \le \sum_{\ell=2}^{m}\mathrm{P}\Big\{\binom{m}{\ell}|H_n^{(\ell)}(\cdot;\mathrm{P}_{jk})| \ge t_2 \frac{4b_1^{2/\ell}\binom{m}{\ell}/C''_\ell}{\sum_{\ell=2}^{m} 4b_1^{2/\ell}\binom{m}{\ell}/C''_\ell}\Big\}$$
$$\le \Big(\sum_{\ell=2}^{m} C'_\ell\Big)\exp\Big\{ - \frac{n t_2}{\sum_{\ell=2}^{m} 4b_1^{2/\ell}\binom{m}{\ell}/C''_\ell}\Big\}.$$

Taking $t_2 = C_2(\log p/n)$, where C_2 is another constant to be specified later, we have

$$\mathrm{P}\Big\{\Big|\sum_{\ell=2}^{k}\binom{m}{\ell} H_n^{(\ell)}(\cdot;\mathrm{P}_{jk})\Big| \ge C_2(\log p/n)\Big\} \le \Big(\sum_{\ell=2}^{m} C'_\ell\Big)\Big(\frac{1}{p}\Big)^{C_2/\{\sum_{\ell=2}^{m} 4b_1^{2/\ell}\binom{m}{\ell}/C''_\ell\}}. \tag{A.1.36}$$

Putting (A.1.35) and (A.1.36) together, and choosing

$$C_1 = 2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} \quad \text{and} \quad C_2 = 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell},$$

we deduce

$$\mathrm{P}\Big[|\widehat{U}_{jk} - U_{jk}| \ge \Big\{2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} + 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell}\Big\}\frac{\log p}{n}\Big] \le \Big(2 + \sum_{\ell=2}^{m} C'_\ell\Big)\frac{1}{p^3}.$$

Then using Slutsky's argument gives

$$\mathrm{P}\Big[\max_{j<k}|\widehat{U}_{jk} - U_{jk}| \ge \Big\{2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} + 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell}\Big\}\frac{\log p}{n}\Big] \le \frac{2 + \sum_{\ell=2}^{m} C'_\ell}{2}\cdot\frac{1}{p},$$

which implies that, with probability at least $1 - (1 + \sum_{\ell=2}^{m} C'_\ell/2)p^{-1}$,

$$\max_{j<k}|\widehat{U}_{jk} - U_{jk}| \le \Big\{2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} + 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell}\Big\}\frac{\log p}{n}.$$

Hence for $n \geq 2$, we have with probability no smaller than $1 - (1 + \sum_{\ell=2}^{m} C'_\ell/2)p^{-1}$,

$$\max_{j<k} \widehat{U}_{jk} \geq \max_{j<k} U_{jk} - \max_{j<k} |\widehat{U}_{jk} - U_{jk}|$$
$$\geq \Big\{ C_\gamma - 2mb_1 - m(4b_1^2 + 6\gamma C_\gamma)^{1/2} - 12 \sum_{\ell=2}^{m} \frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell} \Big\} \frac{\log p}{n} \geq \frac{5\lambda_1\binom{m}{2}\log p}{n-1},$$

where the last inequality is satisfied by choosing C_γ large enough. Accordingly, for any given Q_α, the probability that

$$\frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} \geq 5\log p > 4\log p + (\mu_1 - 2)\log\log p - \frac{\Lambda}{\lambda_1} + Q_\alpha$$

tends to 1 as p goes to infinity. The proof is thus completed. □

A.1.2.6 Proof of Theorem 2.4.4

Proof of Theorem 2.4.4. In view of Corollary 2.4.2, the results follow from Lemma 2.4.1 and the fact that $D_{jk}, R_{jk}, \tau^*_{jk} \asymp \Sigma^2_{jk}$ as $\Sigma_{jk} \to 0$, which has been shown in the proof of Lemma 2.4.1. □

A.1.3 Proofs for Section 2.6 of the main paper

A.1.3.1 Proof of Theorem 2.6.1

Proof of Theorem 2.6.1. The proof of Theorem 2.6.1 hinges on the identity (2.6.1), the fact that random vectors of continuous margins almost surely have no ties among the values of each coordinate, and that $\mathrm{E}h_D \geq 0$ and $\mathrm{E}h_R \geq 0$ (see Hoeffding (1948, p. 547) and Blum et al. (1961, p. 490)).

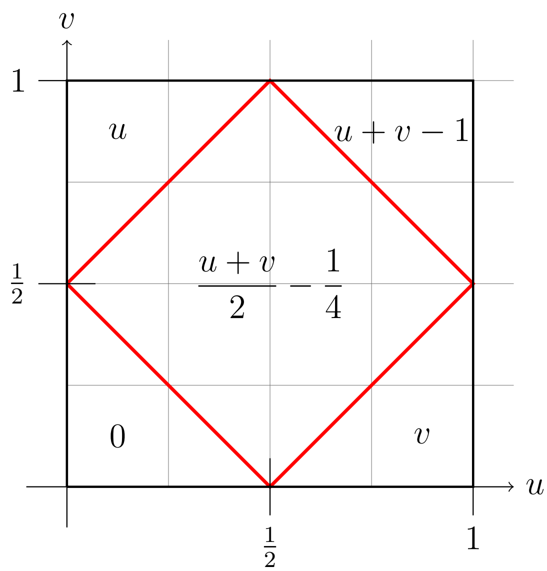
The identity (2.6.1) now gives that $\mathrm{E}h_{\tau^*} \geq 0$ and that $\mathrm{E}h_{\tau^*} = 0$ if and only if $\mathrm{E}h_D = \mathrm{E}h_R = 0$, which in turn implies independence of the considered pair of random variables. □

A.1.3.2 Proof of Proposition 2.6.1

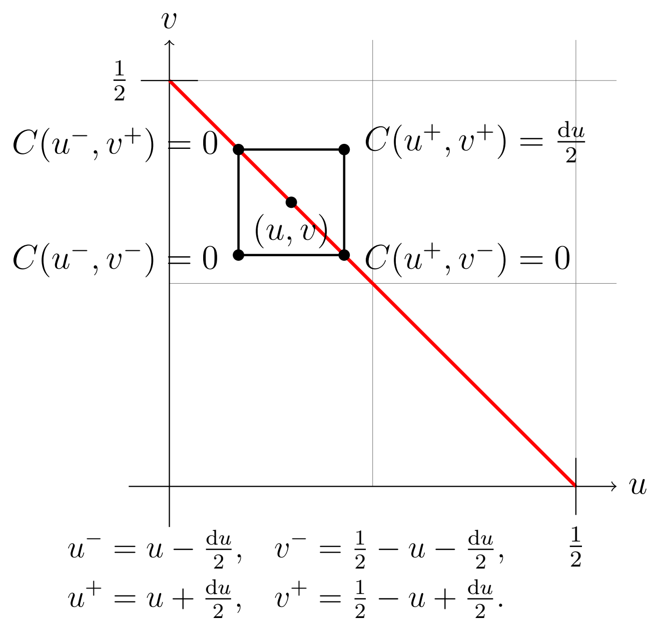
Proof of Proposition 2.6.1. The copula of $(X, Y)^\top$ is given by Nelsen (2006, p. 56):

$$C(u, v) = \begin{cases} \min(u, v), & \text{if } |u - v| \geq \frac{1}{2}, \\ \max(0, u + v - 1), & \text{if } |u + v - 1| \geq \frac{1}{2}, \\ \frac{u+v}{2} - \frac{1}{4}, & \text{otherwise.} \end{cases}$$

We summarize the copula in Figure A.1a.



(a) The copula of the circular uniform distribution with its support marked in red.



(b) Integral on part of the support.

Figure A.1: The copula of the circular uniform distribution.

Since both X and Y are continuous, by the arguments in Schweizer and Wolff (1981), we obtain

$$\begin{aligned}
\mathrm{E}h_D &= 30\int\{F(x,y)-F_1(x)F_2(y)\}^2\mathrm{d}F(x,y)\\
&= 30\int\{C(u,v)-uv\}^2\mathrm{d}C(u,v)\\
\text{and}\quad \mathrm{E}h_R &= 90\int\{F(x,y)-F_1(x)F_2(y)\}^2\mathrm{d}F_1(x)\mathrm{d}F_2(y)\\
&= 90\int\{C(u,v)-uv\}^2\mathrm{d}u\mathrm{d}v.
\end{aligned}$$

We first compute $\mathrm{E}h_D$. Notice that $\partial^2 C(u,v)/\partial u\partial v = 0$ in $[0,1]\times[0,1]$ except for the support of $C(u,v)$ (marked in red in Figure A.1a). Therefore, we only need to compute the integral on the support consisting of four line segments. In Figure A.1b, we illustrate how to find $dC(u,v)$ on the line segment from $(0,1/2)$ to $(1/2,0)$ (denoted by $\mathcal{C}_1$). We have

$$\mathrm{d}C(u,v) = C(u^+,v^+) - C(u^+,v^-) - C(u^-,v^+) + C(u^-,v^-) = \frac{\mathrm{d}u}{2},$$

and thus the integral on the line segment $\mathcal{C}_1$ is given by

$$30\int_{\mathcal{C}_1}\{C(u,v)-uv\}^2\mathrm{d}C(u,v) = 30\int_0^{1/2}\Big\{0-u\Big(\frac{1}{2}-u\Big)\Big\}^2\frac{\mathrm{d}u}{2} = \frac{1}{64}.$$

We can evaluate the integral on the other three line segments (denoted by $\mathcal{C}_2,\mathcal{C}_3,\mathcal{C}_4$, respectively) similarly, and we find

$$\mathrm{E}h_D = 30\int_{\mathcal{C}_1+\mathcal{C}_2+\mathcal{C}_3+\mathcal{C}_4}\{C(u,v)-uv\}^2\mathrm{d}C(u,v) = \frac{1}{16}.$$

The computation of $\mathrm{E}h_R = 90\int\{C(u,v)-uv\}^2\mathrm{d}u\mathrm{d}v = 1/16$ is standard, and we omit details. Finally, using the identity (2.6.1), we deduce that $\mathrm{E}h_{\tau^*} = 1/16$. □

A.2 More comments on τ^*

First of all, we show that the identity (2.6.1) in the main paper may be false when ties exist.

Example A.2.1. If we take $\boldsymbol{z}_i = (\lfloor (i+2)/3 \rfloor, i)^\top$ for $i \in [6]$, then

$$\binom{6}{5}^{-1} \sum_{1 \le i_1 < \cdots < i_5 \le 6} h_D(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_5}) = 1/2, \quad h_R(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6) = 3/2,$$

$$\text{and} \quad \binom{6}{4}^{-1} \sum_{1 \le i_1 < \cdots < i_4 \le 6} h_{\tau^*}(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_4}) = 3/5.$$

In view of Example A.2.1, the proof of Theorem 2.6.1 cannot be directly extended to pairs consisting of both discrete and continuous random variables, and the question if Bergsma–Dassios's conjecture is correct remains open in that regard. However, by the Lebesgue decomposition theorem, in order to prove Bergsma–Dassios's conjecture it suffices to prove the case where the pair follows a mixture of discrete and singular measures.

We now provide a second proof of Theorem 2.6.1 for the absolute continuity case only. It connects the correlation measures raised by Bergsma and Dassios (2014) and the one in the proof of Proposition 9 in Yanagimoto (1970). We believe the resulting alternative representation of the population τ^* is of independent interest, e.g., from the point of view of multivariate extensions of τ^* as considered by Weihs et al. (2018).

Proposition A.2.1. *For any pair of absolutely continuous random variables $(X, Y)^\top \in \mathbb{R}^2$ with joint distribution function $F(x, y)$ and marginal distribution functions $F_1(x), F_2(y)$, we have*

$$\begin{aligned}
&\frac{1}{18}\,\mathrm{E}h_{\tau^*} \\
&\overset{(i)}{=} \int F^2 \mathrm{d}(F + F_1F_2) - \int F^2 \mathrm{d}(FF_1) - 2\int FF_1 \mathrm{d}(FF_2) + \int FF_1 \mathrm{d}(F^2) + \frac{1}{18} \\
&\overset{(ii)}{=} \int F^2 \mathrm{d}F - 2\int FF_1F_2 \mathrm{d}F + 2\int F^2 \mathrm{d}F_1 \mathrm{d}F_2 - \frac{1}{9} \\
&\overset{(iii)}{=} \int (F - F_1F_2)^2 \mathrm{d}F + 2\int (F - F_1F_2)^2 \mathrm{d}F_1 \mathrm{d}F_2 \\
&= \frac{1}{30}\,\mathrm{E}h_D + \frac{1}{45}\,\mathrm{E}h_R,
\end{aligned}$$

where the term on the righthand side of the identity (ii) is Yanagimoto's correlation measure.

Proof of Proposition A.2.1. We prove identities (i)–(iii) sequentially. Let Ψ_1, Ψ_2, Ψ_3 denote the expressions on the right-hand side of identities (i), (ii), (iii), respectively.

Identity (i)**.** Let $\{(X_i, Y_i)^\top\}_{i\in[4]}$ be four independent realizations of $(X, Y)^\top$. For Bergsma–Dassios–Yanagimoto's τ^*, we have, by Equation (6) in Bergsma and Dassios (2014),

$$\begin{aligned}\frac{1}{18}\,\mathrm{E}h_{\tau^*} = &\ \frac{1}{3}\,\mathrm{P}\{\max(X_1,X_2) < \min(X_3,X_4),\ \ \max(Y_1,Y_2) < \min(Y_3,Y_4)\}\\ &+\frac{1}{3}\,\mathrm{P}\{\max(X_1,X_2) < \min(X_3,X_4),\ \ \max(Y_3,Y_4) < \min(Y_1,Y_2)\}\\ &-\frac{2}{3}\,\mathrm{P}\{\max(X_1,X_2) < \min(X_3,X_4),\ \ \max(Y_1,Y_3) < \min(Y_2,Y_4)\}.\end{aligned} \tag{A.2.1}$$

We study the three terms in (A.2.1) separately, starting from the first term. Using Fubini's theorem, we get

$$\begin{aligned}&\mathrm{P}\{\max(X_1,X_2) < \min(X_3,X_4),\ \ \max(Y_1,Y_2\} < \min(Y_3,Y_4)\}\\ =&\int \mathrm{P}\{\max(X_1,X_2) < x,\ \ \max(Y_1,Y_2) < y\} d\mathrm{P}\{\min(X_3,X_4)\le x,\ \ \min(Y_3,Y_4)\le y\}\\ =&\int F(x,y)^2 \mathrm{dP}\{\min(X_3,X_4)\le x,\ \ \min(Y_3,Y_4)\le y\},\end{aligned} \tag{A.2.2}$$

where

$$\mathrm{P}\{\min(X_3,X_4)\le x,\ \ \min(Y_3,Y_4)\le y\} = \mathrm{P}\{(A\cup B)\cap(C\cup D)\} = \mathrm{P}(I\cup II\cup III\cup IV)$$

and $A := \{X_3 \le x\}$, $B := \{X_4 \le X\}$, $C := \{Y_3 \le y\}$, $D := \{Y_4 \le y\}$,

$$\begin{aligned}I &:= A\cap C = \{X_3\le x, Y_3\le y\}, &\quad II &:= A\cap D = \{X_3\le x, Y_4\le y\},\\ III &:= B\cap C = \{X_4\le x, Y_3\le y\}, &\quad IV &:= B\cap D = \{X_4\le x, Y_4\le y\}.\end{aligned}$$

From the inclusion–exclusion principle, we obtain

$$
\begin{aligned}
&\mathrm{P}\{\min(X_3, X_4) \le x,\ \ \min(Y_3, Y_4) \le y\} \\
&= \mathrm{P}(I) + \mathrm{P}(II) + \mathrm{P}(III) + \mathrm{P}(IV) \\
&\quad - \mathrm{P}(I \cap II) - \mathrm{P}(I \cap III) - \mathrm{P}(I \cap IV) - \mathrm{P}(II \cap III) - \mathrm{P}(II \cap IV) - \mathrm{P}(III \cap IV) \\
&\quad + \mathrm{P}(I \cap II \cap III) + \mathrm{P}(I \cap II \cap IV) + \mathrm{P}(I \cap III \cap IV) + \mathrm{P}(II \cap III \cap IV) \\
&\quad - \mathrm{P}(I \cap II \cap III \cap IV) \\
&= F + F_1F_2 + F_1F_2 + F - FF_2 - FF_1 - F^2 - F^2 - FF_1 - FF_2 + F^2 + F^2 + F^2 + F^2 - F^2 \\
&= 2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2. \qquad \text{(A.2.3)}
\end{aligned}
$$

Plugging (A.2.3) into (A.2.2) implies that

$$
\begin{aligned}
&\mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2\} < \min(Y_3, Y_4)\} \\
&= \int F^2 \mathrm{d}(2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2). \qquad \text{(A.2.4)}
\end{aligned}
$$

The second term in (A.2.1) can be written as

$$
\begin{aligned}
&\mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\} \\
&= \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4)\} \\
&\quad - \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \min(Y_1, Y_2) \le \max(Y_3, Y_4)\} \\
&= \int \mathrm{P}\{\max(X_1, X_2) < x\} \mathrm{dP}\{\min(X_3, X_4) \le x\} \\
&\quad - \mathrm{P}\{\max(X_1, X_2) < x,\ \ \min(Y_1, Y_2) \le y\} \mathrm{dP}\{\min(X_3, X_4) \le x,\ \ \max(Y_3, Y_4) \le y\}, \\
&\qquad \text{(A.2.5)}
\end{aligned}
$$

where we have

$$
\mathrm{P}\{\min(X_3, X_4) \le x\} = \mathrm{P}(A \cup B) = 2F_1 - F_1^2, \qquad \text{(A.2.6)}
$$

and

$$\begin{aligned}
&\mathrm{P}\{\max(X_1,X_2)<x,\ \ \min(Y_1,Y_2)\le y\}\\
&=\mathrm{P}\{\max(X_1,X_2)<x,\ Y_1\le y\}+\mathrm{P}\{\max(X_1,X_2)<x,\ Y_2\le y\}\\
&\quad-\mathrm{P}\{\max(X_1,X_2)<x,\ \ \max(Y_1,Y_2)\le y\}\\
&=2FF_1-F^2,
\end{aligned}\tag{A.2.7}$$

and

$$\begin{aligned}
&\mathrm{P}(\min\{X_3,X_4\}\le x,\ \ \max\{Y_3,Y_4\}\le y)\\
&=\mathrm{P}[\{(X_3\le x)\cup(X_4\le x)\}\cap\{(Y_3\le y)\cap(Y_4\le y)\}]\\
&=\mathrm{P}[\{A\cap(C\cap D)\}\cup\{B\cap(C\cap D)\}]\\
&=2FF_2-F^2.
\end{aligned}\tag{A.2.8}$$

Plugging (A.2.6)–(A.2.8) into (A.2.5) yields

$$\begin{aligned}
&\mathrm{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \ \max(Y_3,Y_4)<\min(Y_1,Y_2)\}\\
&=\int F_1^2\mathrm{d}(2FF_1-F^2)-\int(2FF_1-F^2)\mathrm{d}(2FF_2-F^2).
\end{aligned}\tag{A.2.9}$$

Next we handle the third term in (A.2.1). We have by symmetry that

$$\begin{aligned}
&\mathrm{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \ \max(Y_1,Y_3)<\min(Y_2,Y_4)\}\\
&=\mathrm{P}\{\max(X_1,X_2)<\min(X_4,X_3),\ \ \max(Y_1,Y_4)<\min(Y_2,Y_3)\}\\
&=\mathrm{P}\{\max(X_2,X_1)<\min(X_3,X_4),\ \ \max(Y_2,Y_3)<\min(Y_1,Y_4)\}\\
&=\mathrm{P}\{\max(X_2,X_1)<\min(X_4,X_3),\ \ \max(Y_2,Y_4)<\min(Y_1,Y_3)\}.
\end{aligned}\tag{A.2.10}$$

We also notice that

$$\mathrm{P}\{\max(X_1,X_2)<\min(X_3,X_4)$$

$$
\begin{aligned}
&= \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2) < \min(Y_3, Y_4)\} \\
&\quad + \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\} \\
&\quad + \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_3) < \min(Y_2, Y_4)\} \\
&\quad + \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_4) < \min(Y_2, Y_3)\} \\
&\quad + \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_2, Y_3) < \min(Y_1, Y_4)\} \\
&\quad + \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_2, Y_4) < \min(Y_1, Y_3)\} \qquad \text{(A.2.11)}
\end{aligned}
$$

assuming marginal continuity of $(X, Y)^\top$. Combining (A.2.10) and (A.2.11) gives

$$
\begin{aligned}
&\mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_3) < \min(Y_2, Y_4)\} \\
&= \frac{1}{4}\Big[\mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4) \\
&\quad - \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2) < \min(Y_3, Y_4)\} \\
&\quad - \mathrm{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\}\Big] \\
&= \frac{1}{4}\Big\{\int (2FF_1 - F^2)\mathrm{d}(2FF_2 - F^2) - \int F^2 \mathrm{d}(2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2)\Big\}.
\end{aligned}
\qquad \text{(A.2.12)}
$$

The identity (i) follows by plugging (A.2.4), (A.2.9), and (A.2.12) into (A.2.1).

Identity (ii)**.** Next we prove that $\Psi_1 - \Psi_2 = 0$. A straightforward computation gives

$$
\begin{aligned}
&\Psi_1 - \Psi_2 \\
&= -\int F^2 d(F_1F_2) - \int F^2 \mathrm{d}(FF_1) - 2\int FF_1 \mathrm{d}(FF_2) + \int FF_1 \mathrm{d}(F^2) + 2\int FF_1F_2 \mathrm{d}F + \frac{1}{6} \\
&= -\iint F^2 \frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y - \iint F^2 \frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}\mathrm{d}x\mathrm{d}y + \iint F^2 F_1 \frac{\partial^2 F}{\partial x \partial y}\mathrm{d}x\mathrm{d}y \\
&\quad - 2\iint FF_1 \frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y + 2\iint FF_1 \frac{\partial F}{\partial x}\frac{\partial F}{\partial y}\mathrm{d}x\mathrm{d}y + \frac{1}{6}.
\end{aligned}
\qquad \text{(A.2.13)}
$$

To further simplify (A.2.13), notice that

$$\iint \Big(F^2\frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}+2FF_1\frac{\partial F}{\partial x}\frac{\partial F}{\partial y}+F^2F_1\frac{\partial^2 F}{\partial x\partial y}\Big)\mathrm{d}x\mathrm{d}y=\iint \frac{\partial^2(F^3F_1/3)}{\partial x\partial y}\mathrm{d}x\mathrm{d}y=\frac{1}{3}. \tag{A.2.14}$$

Adding (A.2.13) and (A.2.14) together yields

$$\begin{aligned}
\Psi_1-\Psi_2 &= -\Big(\iint F^2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y+2\iint FF_1\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y\Big)-2\iint F^2\frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}\mathrm{d}x\mathrm{d}y+\frac{1}{2}\\
&= -\int\frac{\partial F_2}{\partial y}\int\frac{\partial(F^2F_1)}{\partial x}\mathrm{d}x\mathrm{d}y-2\int\frac{\partial F_1}{\partial x}\int F^2\frac{\partial F}{\partial y}\mathrm{d}y\mathrm{d}x+\frac{1}{2}\\
&= -\int\frac{\partial F_2}{\partial y}F_2^2\mathrm{d}y-2\int\frac{\partial F_1}{\partial x}\frac{F_1^3}{3}\mathrm{d}x+\frac{1}{2}\\
&= -\frac{1}{3}-2\Big(\frac{1}{12}\Big)+\frac{1}{2}=0,
\end{aligned}$$

which completes the proof of identity (*ii*).

Identity (*iii*). This identity was discovered by Yanagimoto (1970). To see this, it suffices to show that

$$\Psi_3-\Psi_2=\int F_1^2F_2^2dF-4\int FF_1F_2dF_1dF_2+2\int F_1^2F_2^2dF_1dF_2+\frac{1}{9}=0.$$

We start from the identity

$$\begin{aligned}
1=\iint\frac{\partial^2(FF_1^2F_2^2)}{\partial x\partial y}\mathrm{d}x\mathrm{d}y=&\iint F_1^2F_2^2\frac{\partial^2F}{\partial x\partial y}\mathrm{d}x\mathrm{d}y+\iint 4FF_1F_2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y\\
&+\iint 2F_1^2F_2\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y+\iint 2F_1F_2^2\frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}\mathrm{d}x\mathrm{d}y.
\end{aligned} \tag{A.2.15}$$

We also note that

$$\iint 2FF_1F_2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y+\iint F_1^2F_2\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}\mathrm{d}x\mathrm{d}y$$

$$= \int F_2 \frac{\partial F_2}{\partial y} \int \frac{\partial (F F_1^2)}{\partial x} \mathrm{d}x \mathrm{d}y = \int F_2^2 \frac{\partial F_2}{\partial y} \mathrm{d}y = \frac{1}{3}, \tag{A.2.16}$$

and

$$\begin{aligned} &\iint 2F F_1 F_2 \frac{\partial F_1}{\partial x} \frac{\partial F_2}{\partial y} \mathrm{d}x \mathrm{d}y + \iint F_1 F_2^2 \frac{\partial F_1}{\partial x} \frac{\partial F}{\partial y} \mathrm{d}x \mathrm{d}y \\ &= \int F_1 \frac{\partial F_1}{\partial x} \int \frac{\partial (F F_2^2)}{\partial y} \mathrm{d}y \mathrm{d}x = \int F_1^2 \frac{\partial F_1}{\partial x} \mathrm{d}x = \frac{1}{3}, \end{aligned} \tag{A.2.17}$$

and

$$\iint 2F_1^2 F_2^2 \frac{\partial F_1}{\partial x} \frac{\partial F_2}{\partial y} \mathrm{d}x \mathrm{d}y = 2 \int F_1^2 \frac{\partial F_1}{\partial x} \mathrm{d}x \int F_2^2 \frac{\partial F_2}{\partial y} \mathrm{d}y = 2\Big(\frac{1}{3}\Big)\Big(\frac{1}{3}\Big) = \frac{2}{9}. \tag{A.2.18}$$

Combining (A.2.15)–(A.2.18) concludes the claim. □

A.3 Additional simulation results

First, we report the sizes and powers of the proposed tests with simulation-based critical values ($M = 5,000$) as shown in Table A.1. The table shows results only for Examples 2.5.1, 2.5.3, and 2.5.4 as the simulated powers under Example 2.5.2 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.

Next, in order to interpret the power in Examples 2.5.2–2.5.4, we consider the following example.

Example 2.5.2–2.5.4 (continued). *We consider modified data drawn as*

$$\boldsymbol{X}_\alpha = \alpha \boldsymbol{X} + (1 - \alpha)\boldsymbol{\mathcal{E}}$$

where $\alpha \in [0, 1]$ *represents the level of a desired signal,* $\boldsymbol{X}$ *is the same as that in Examples 2.5.2–2.5.4, respectively, and* $\boldsymbol{\mathcal{E}} \sim N_p(0, \mathbf{I}_p)$ *is independent of* $\boldsymbol{X}$*.*

Table A.1: Empirical sizes and powers of simulation-based rejection threshold in Examples 2.5.1–2.5.4 (The powers under Example 2.5.2 are all perfectly 1.000 and hence omitted)

n	p	DHS_D	DHS_R	DHS_{τ^*}	DHS_D	DHS_R	DHS_{τ^*}	DHS_D	DHS_R	DHS_{τ^*}	DHS_D	DHS_R	DHS_{τ^*}
		Example 2.5.1(a)			Example 2.5.3(a)			Example 2.5.3(b)			Example 2.5.3(c)		
100	50	0.053	0.053	0.053	0.964	0.964	0.965	0.746	0.651	0.694	0.639	0.591	0.611
	100	0.051	0.051	0.050	0.955	0.954	0.955	0.731	0.636	0.676	0.638	0.581	0.607
	200	0.045	0.045	0.044	0.943	0.944	0.945	0.698	0.602	0.643	0.609	0.549	0.580
	400	0.045	0.046	0.046	0.930	0.931	0.932	0.674	0.577	0.624	0.592	0.524	0.557
	800	0.054	0.051	0.051	0.921	0.921	0.923	0.651	0.548	0.594	0.567	0.490	0.526
200	50	0.050	0.053	0.051	0.991	0.991	0.991	0.896	0.853	0.872	0.822	0.800	0.810
	100	0.048	0.048	0.047	0.984	0.985	0.985	0.874	0.824	0.847	0.803	0.775	0.787
	200	0.046	0.045	0.044	0.983	0.984	0.984	0.852	0.794	0.820	0.785	0.757	0.769
	400	0.051	0.058	0.055	0.983	0.984	0.984	0.842	0.778	0.805	0.766	0.738	0.751
	800	0.042	0.044	0.046	0.978	0.978	0.979	0.809	0.746	0.776	0.741	0.708	0.727
		Example 2.5.4(a)			Example 2.5.4(b)			Example 2.5.4(c)					
100	50	0.081	0.085	0.080	0.096	0.094	0.096	0.121	0.124	0.126			
	100	0.079	0.074	0.077	0.074	0.074	0.074	0.088	0.090	0.092			
	200	0.052	0.059	0.056	0.067	0.069	0.068	0.072	0.072	0.074			
	400	0.064	0.064	0.064	0.059	0.057	0.056	0.059	0.058	0.065			
	800	0.051	0.048	0.048	0.058	0.054	0.052	0.061	0.064	0.059			
200	50	0.099	0.099	0.098	0.110	0.114	0.112	0.115	0.120	0.115			
	100	0.060	0.064	0.063	0.081	0.084	0.080	0.090	0.091	0.087			
	200	0.066	0.067	0.071	0.046	0.046	0.044	0.080	0.070	0.079			
	400	0.058	0.062	0.058	0.060	0.070	0.069	0.059	0.058	0.058			
	800	0.045	0.049	0.050	0.052	0.050	0.050	0.061	0.060	0.062			

The relationships between empirical powers (5,000 replicates) based on observations from $\boldsymbol{X}_\alpha$ and the value α for Examples 2.5.2–2.5.4 (continued) are summarized in Figures A.2–A.5. As expected, the power of each test is monotonically increasing in α, i.e., as the signal increases. Similar patterns as we discussed for Examples 2.5.2–2.5.4 can be found here. It can be noticed that, the three proposed tests, followed by LD_{τ^*}, uniformly dominate the other tests in Examples 2.5.2 and 2.5.3 (continued) that are sparse settings.

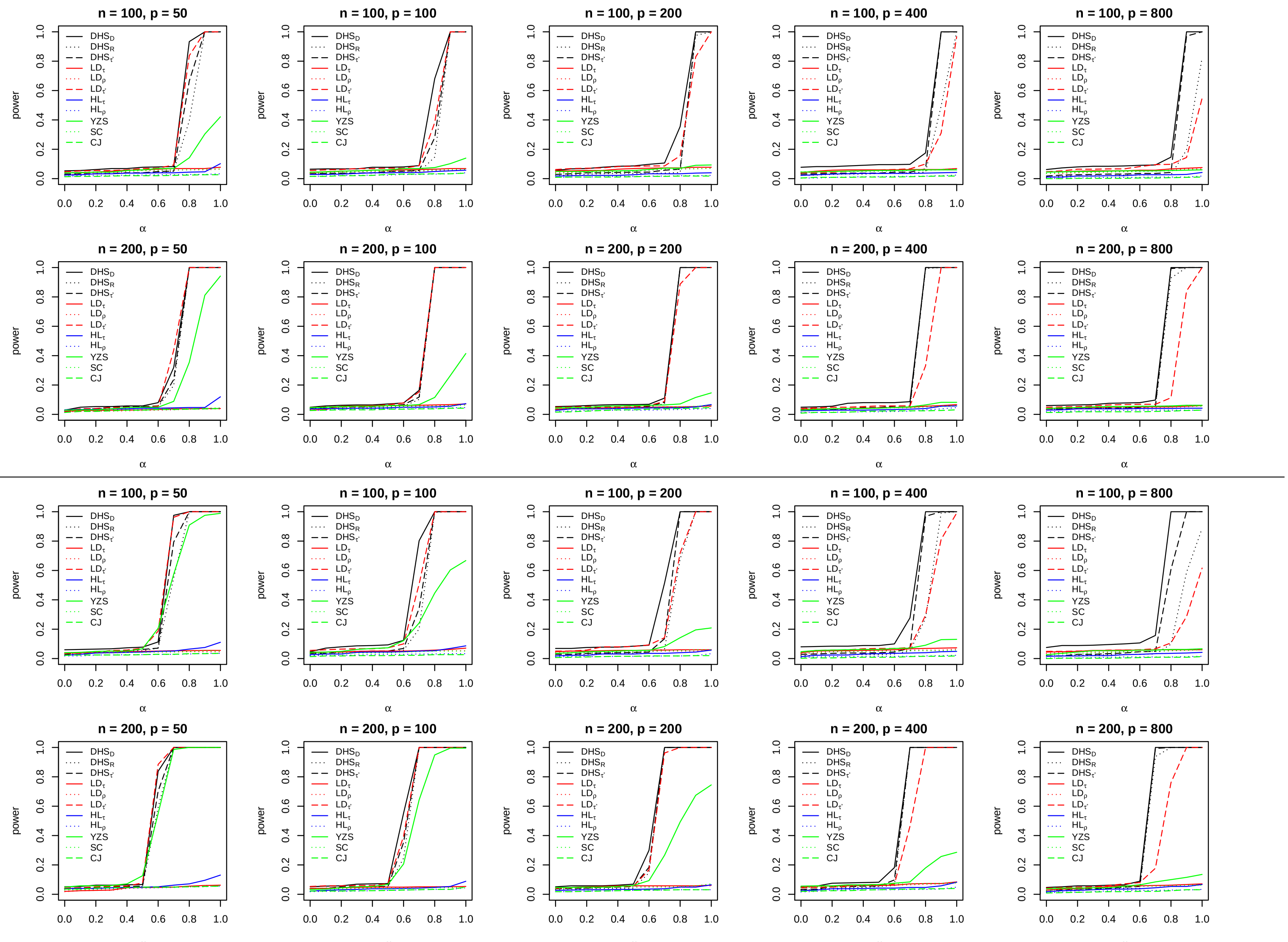


Figure A.2: Empirical powers of the eleven competing tests in continued Example 2.5.2(a) (first two rows) and continued Example 2.5.2(b) (last two rows). The y-axis represents the power based on 5,000 replicates and the x-axis represents the level of a desired signal.

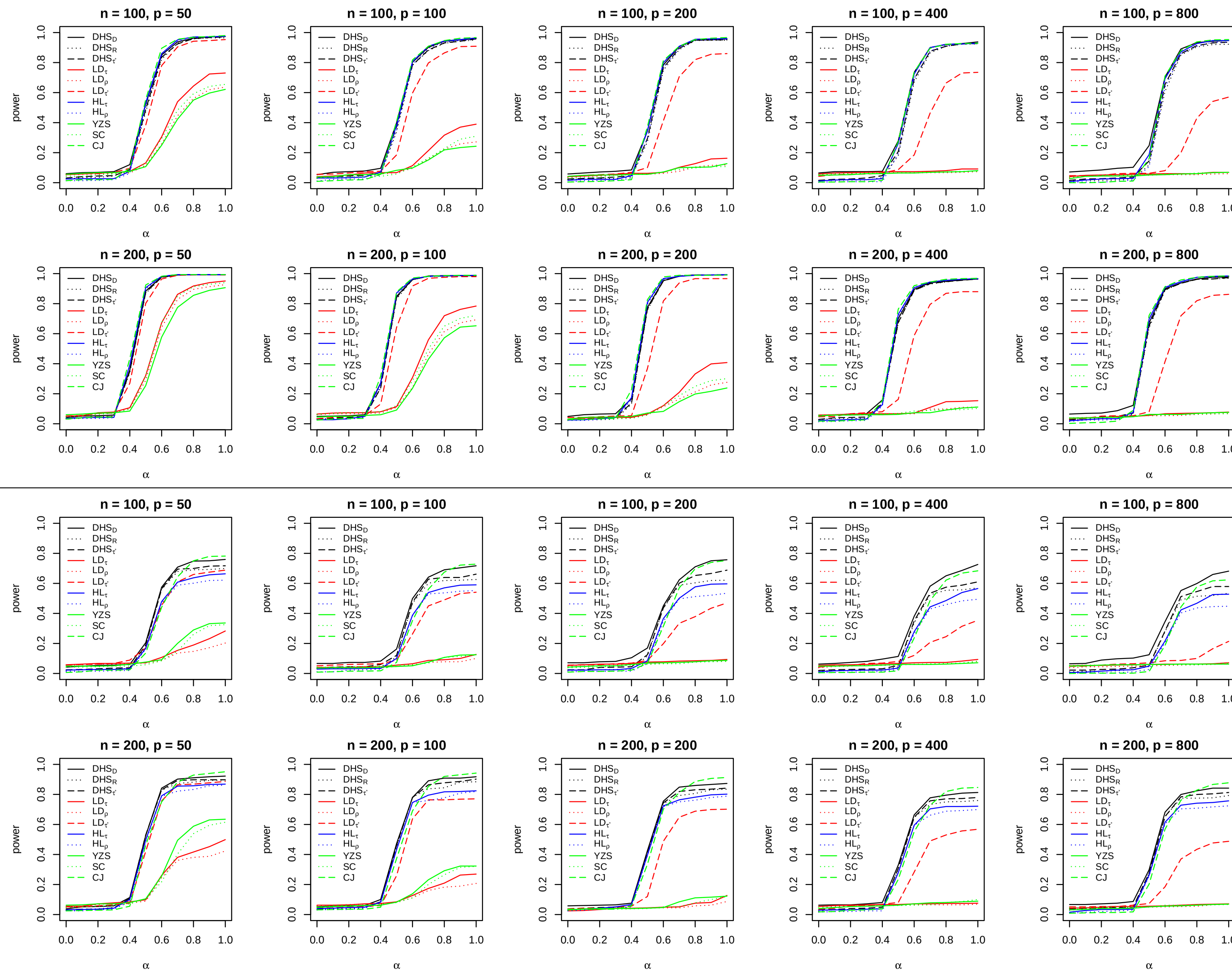


Figure A.3: Empirical powers of the eleven competing tests in continued Example 2.5.3(a) (first two rows) and continued Example 2.5.3(b) (last two rows). The y-axis represents the power based on 5,000 replicates and the x-axis represents the level of a desired signal.

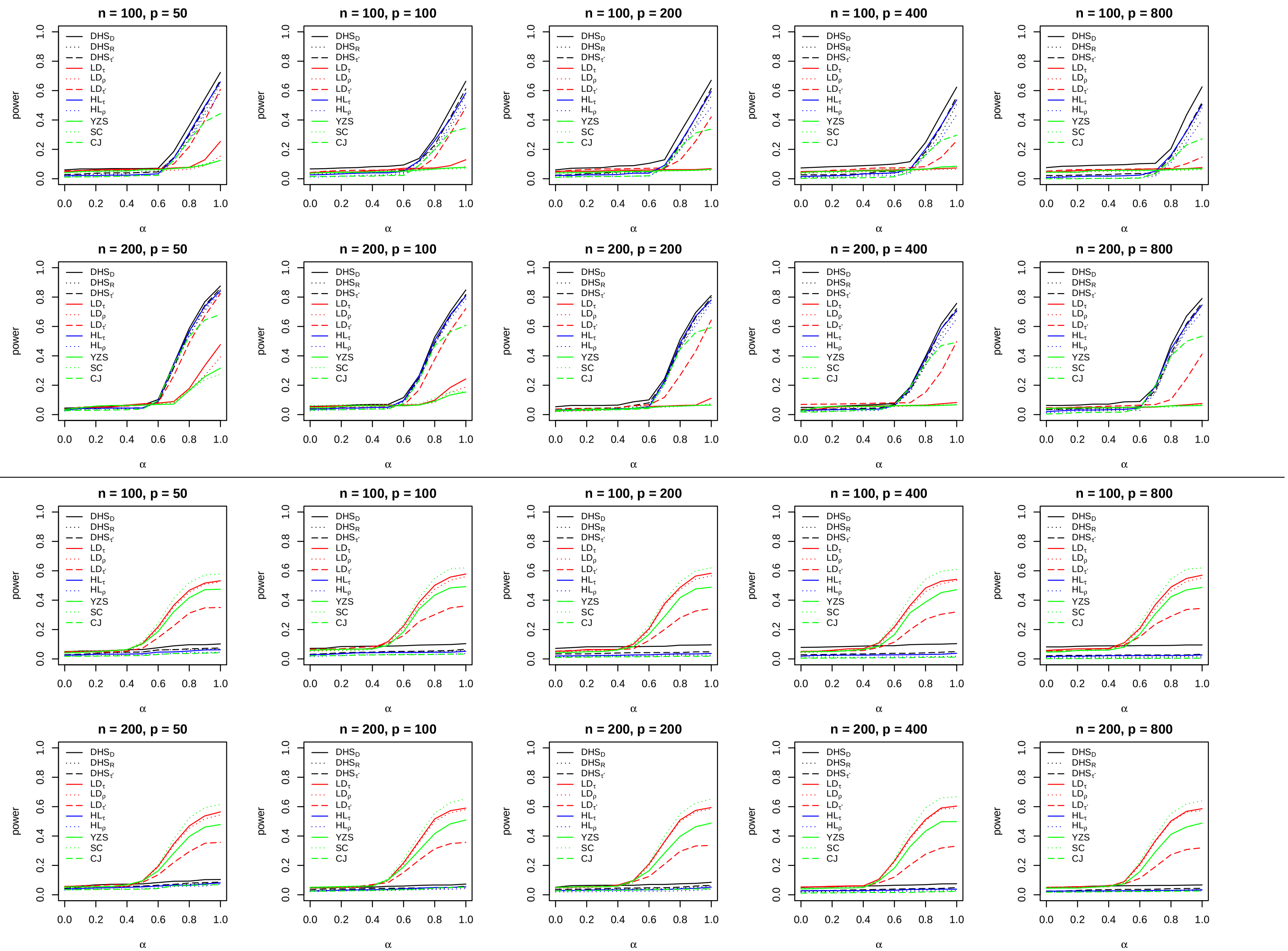


Figure A.4: Empirical powers of the eleven competing tests in continued Example 2.5.3(c) (first two rows) and continued Example 2.5.4(a) (last two rows). The y-axis represents the power based on 5,000 replicates and the x-axis represents the level of a desired signal.

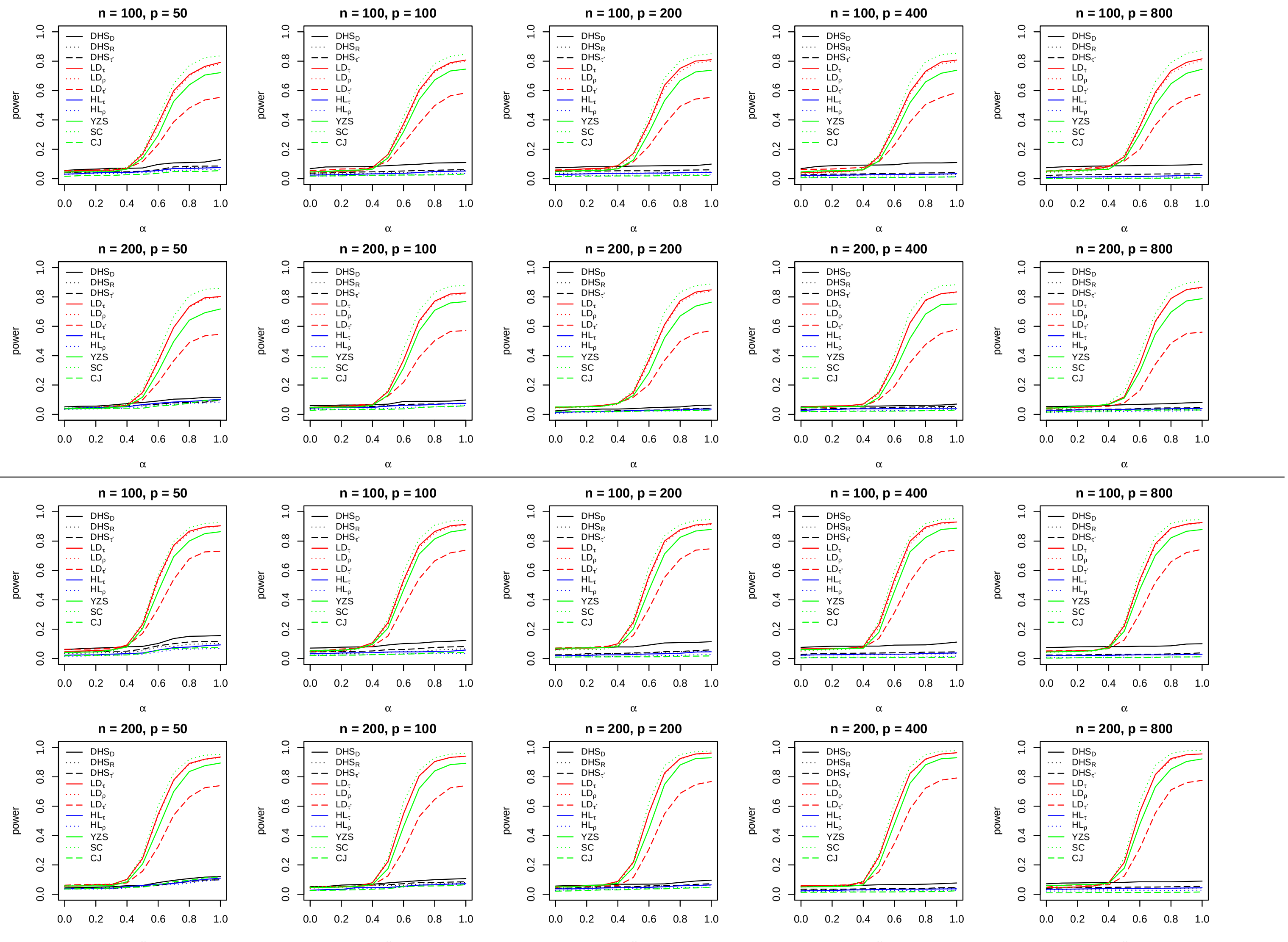


Figure A.5: Empirical powers of the eleven competing tests in continued Example 2.5.4(b) (first two rows) and continued Example 2.5.4(c) (last two rows). The y-axis represents the power based on 5,000 replicates and the x-axis represents the level of a desired signal.

Appendix B
APPENDIX OF CHAPTER 3

B.1 Proofs

Some further concepts and notation concerning U-statistics are needed in this section. For any symmetric kernel h, any integer $\ell \in [\![m]\!]$, and any probability measure $\mathrm{P}_{\boldsymbol{Z}}$, recall the definition

$$h_\ell(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) := \mathrm{E}h(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell, \boldsymbol{Z}_{\ell+1}, \ldots, \boldsymbol{Z}_m),$$

of the kernel and define

$$\widetilde{h}_\ell(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) := h_\ell(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}}) - \mathrm{E}h - \sum_{k=1}^{\ell-1} \sum_{1 \le i_1 < \cdots < i_k \le \ell} \widetilde{h}_k(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_k}; \mathrm{P}_{\boldsymbol{Z}}),$$

where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are m independent copies of $\boldsymbol{Z} \sim \mathrm{P}_{\boldsymbol{Z}}$ and $\mathrm{E}h := \mathrm{E}h(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m)$. The kernel as well as the corresponding U-statistic are said to be *degenerate* under $\mathrm{P}_{\boldsymbol{Z}}$ if $h_1(\cdot)$ has variance zero and *completely degenerate* if the variances of $h_1(\boldsymbol{Z}_1), \ldots, h_{m-1}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m)$ all are zero. We also have, for any (possibly dependent) random vectors $\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_n'$,

$$\binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h\Big(\boldsymbol{Z}_{i_1}', \ldots, \boldsymbol{Z}_{i_m}'\Big) = \mathrm{E}h + \sum_{\ell=1}^{m} \binom{n}{\ell}^{-1} \sum_{1 \le i_1 < \cdots < i_\ell \le n} \binom{m}{\ell} \widetilde{h}_\ell\Big(\boldsymbol{Z}_{i_1}', \ldots, \boldsymbol{Z}_{i_\ell}'; \mathrm{P}_{\boldsymbol{Z}}\Big),$$

(the so-called *Hoeffding decomposition* with respect to $\mathrm{P}_{\boldsymbol{Z}}$).

Notation. The cardinality of a set $\mathcal{S}$ is denoted as $\mathrm{card}(\mathcal{S})$ and its complement as $\mathcal{S}^{\complement}$. We use $\rightrightarrows$ to denote uniform convergence of functions The cumulative distribution function and probability density function of the univariate standard normal distribution are denoted by Φ and φ, respectively. Let $\|X\|_{\mathsf{L}^r} := (\mathrm{E}|X|^r)^{1/r}$ stand for the L^r-norm of a random variable X.

We use $\xrightarrow{\mathsf{L^r}}$ to denote convergence of random variables in the r-th mean. For random vectors $\boldsymbol{X}_n, \boldsymbol{X} \in \mathbb{R}^d$, we write $\boldsymbol{X}_n \xrightarrow{\mathsf{L^r}} \boldsymbol{X}$ if $\|\boldsymbol{X}_n - \boldsymbol{X}\| \xrightarrow{\mathsf{L^r}} 0$. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and let P and Q be two probability measures on $(\mathcal{X}, \mathcal{A})$: we write $\mathrm{P} \ll \mu$ and $\mathrm{Q} \ll \mu$ if P and Q are absolutely continuous with respect to a σ-finite measure μ on $(\mathcal{X}, \mathcal{A})$. The total variation and Hellinger distances between Q and P are denoted as $\mathrm{TV}(\mathrm{Q}, \mathrm{P}) := \sup_{A \in \mathcal{A}} |\mathrm{Q}(A) - \mathrm{P}(A)|$ and $\mathrm{HL}(\mathrm{Q}, \mathrm{P}) := \{\int 2(1 - \sqrt{\mathrm{dQ}/\mathrm{dP}})\mathrm{dP}\}^{1/2}$, respectively. We write $\mathrm{Q}^{(n)} \triangleleft \mathrm{P}^{(n)}$ for "$\mathrm{Q}^{(n)}$ is contiguous to $\mathrm{P}^{(n)}$".

B.1.1 Proofs for Section 3.2

B.1.1.1 Proof of Propostion 3.2.1

Proof of Propostion 3.2.1. The proof is entirely similar to the proof of Proposition 2 in Weihs et al. (2018) and hence omitted. □

B.1.1.2 Proof of Example 3.2.1

Proof of Example 3.2.1. Item (a) is stated in Bergsma and Dassios (2014, Sec. 3.4). Item (b) is given in Weihs et al. (2018, Proposition 1). Item (c) can be proved using Equation (3) in Zhu et al. (2017). Items (d) and (e) can be proved using Proposition D.5 in Kim et al. (2020c) and Theorem 7.2 in Kim et al. (2020b), respectively. □

B.1.1.3 Proof of Lemma 3.2.1

Proof of Lemma 3.2.1. Provided that $\mathrm{E}[f_1]$ and $\mathrm{E}[f_2]$ exist and are finite, we have

$$\begin{aligned}
&\mathrm{E}\Big[k_{f_1,f_2,H_*^m}\Big((\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1m}, \boldsymbol{X}_{2m})\Big)\Big] \\
&\quad = \mathrm{E}\Big\{\sum_{\sigma \in H} \mathrm{sgn}(\sigma) f_1(\boldsymbol{X}_{1\sigma(1)}, \ldots, \boldsymbol{X}_{1\sigma(m)})\Big\}\Big\{\sum_{\sigma \in H} \mathrm{sgn}(\sigma) f_2(\boldsymbol{X}_{2\sigma(1)}, \ldots, \boldsymbol{X}_{2\sigma(m)})\Big\} \\
&\quad = \mathrm{E}\Big\{f_1(\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \boldsymbol{X}_{13}, \boldsymbol{X}_{14}, \boldsymbol{X}_{15}, \ldots, \boldsymbol{X}_{1m}) - f_1(\boldsymbol{X}_{11}, \boldsymbol{X}_{13}, \boldsymbol{X}_{12}, \boldsymbol{X}_{14}, \boldsymbol{X}_{15}, \ldots, \boldsymbol{X}_{1m})
\end{aligned}$$

$$- f_1(\boldsymbol{X}_{14}, \boldsymbol{X}_{12}, \boldsymbol{X}_{13}, \boldsymbol{X}_{11}, \boldsymbol{X}_{15}, \dots, \boldsymbol{X}_{1m}) + f_1(\boldsymbol{X}_{14}, \boldsymbol{X}_{13}, \boldsymbol{X}_{12}, \boldsymbol{X}_{11}, \boldsymbol{X}_{15}, \dots, \boldsymbol{X}_{1m})\Big\}$$

$$\times \Big\{ f_2(\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \boldsymbol{X}_{23}, \boldsymbol{X}_{24}, \boldsymbol{X}_{25}, \dots, \boldsymbol{X}_{2m}) - f_2(\boldsymbol{X}_{21}, \boldsymbol{X}_{23}, \boldsymbol{X}_{22}, \boldsymbol{X}_{24}, \boldsymbol{X}_{25}, \dots, \boldsymbol{X}_{2m})$$

$$- f_2(\boldsymbol{X}_{24}, \boldsymbol{X}_{22}, \boldsymbol{X}_{23}, \boldsymbol{X}_{21}, \boldsymbol{X}_{25}, \dots, \boldsymbol{X}_{2m}) + f_2(\boldsymbol{X}_{24}, \boldsymbol{X}_{23}, \boldsymbol{X}_{22}, \boldsymbol{X}_{21}, \boldsymbol{X}_{25}, \dots, \boldsymbol{X}_{2m})\Big\}.$$

The result follows. □

B.1.1.4 Proof of Theorem 3.2.1

Proof of Theorem 3.2.1. The D-consistency of the pairs of kernels used in Example 3.2.1(a) has been shown in Székely et al. (2007, Theorem 3(i)), Lyons (2013, Theorem 3.11) and Lyons (2018, Item (iv)). The result for 3.2.1(b) is given in Weihs et al. (2018, Theorem 1), that for 3.2.1(c) in Zhu et al. (2017, Proposition 1(i)), and that for 3.2.1(d) and 3.2.1(e) in Kim et al. (2020b, p. 3435). □

B.1.1.5 Proof of Lemma 3.2.2

Proof of Lemma 3.2.2. The lemma directly follows from the definition of $\mu_{f_1,f_2,H}$ (cf. Definition 3.2.1) and the fact that f_1 and f_2 are both orthogonally invariant. □

B.1.1.6 Proof of Proposition 3.2.2

Proof of Proposition 3.2.2. To verify that the kernels used in Example 3.2.1(a),(c)–(e) are orthogonally invariant, it suffices to notice that $\mathbf{O}\boldsymbol{w} - \mathbf{O}\boldsymbol{v} = \mathbf{O}(\boldsymbol{w} - \boldsymbol{v})$, $(\mathbf{O}\boldsymbol{w})^\top(\mathbf{O}\boldsymbol{v}) = \boldsymbol{w}^\top\mathbf{O}^\top\mathbf{O}\boldsymbol{v} = \boldsymbol{w}^\top\boldsymbol{v}$, and $\|\mathbf{O}\boldsymbol{w}\| = \sqrt{\boldsymbol{w}^\top\boldsymbol{w}} = \|\boldsymbol{w}\|$ for any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^d$. □

B.1.2 Proofs for Section 3.4

B.1.2.1 Proof of Proposition 3.4.1

Proof of Proposition 3.4.1. The first part is trivial. We next prove the second part. The function $u \mapsto \big(F^{-1}_{\chi^2_d}(u)\big)^{1/2}$ is continuous over $[0, 1)$, and

$$\int_0^1 \Big(\big(F^{-1}_{\chi^2_d}(u)\big)^{1/2}\Big)^2 \mathrm{d}u = \int_0^1 F^{-1}_{\chi^2_d}(u)\mathrm{d}u = \mathrm{E}[F^{-1}_{\chi^2_d}(U)] = d,$$

where U is uniformly distributed over $[0, 1]$, and thus $F^{-1}_{\chi^2_d}(U)$ is chi-square distributed with d degrees of freedom and expectation d. Hence, $J_{\mathrm{vdW}}(u)$ is weakly regular; it is not strongly regular, however, since it is unbounded. □

B.1.2.2 Proof of Proposition 3.4.2

B.1.2.2.1 Proof of Proposition 3.4.2(i)

Proof of Proposition 3.4.2(i). This follows immediately from Proposition B.2.1(iv) and the independence between $[\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i})]^n_{i=1}$ and $[\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i})]^n_{i=1}$ under the null hypothesis. □

B.1.2.2.2 Proof of Proposition 3.4.2(ii)

Proof of Proposition 3.4.2(ii). The desired result follows from combining Lemma 3.2.2 and Proposition B.2.1(iii). □

B.1.2.2.3 Proof of Proposition 3.4.2(iii)

Proof of Proposition 3.4.2(iii). We only prove the D-consistency part. Using Lemma 3.2.1, it remains to prove that the independence of $\mathbf{G}_{1,\pm}(\boldsymbol{X}_1)$ and $\mathbf{G}_{2,\pm}(\boldsymbol{X}_2)$ implies the independence of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Notice that $\mathbf{F}_\pm$ is P-almost surely invertible for any $\mathrm{P} \in \mathcal{P}^{\mathrm{ac}}_d$ (Ambrosio et al., 2008, Section 6.2.3 and Remark 6.2.11), and so is $\mathbf{G}_\pm$. The independence claim follows. □

B.1.2.2.4 Proof of Proposition 3.4.2(iv)

Proof of Proposition 3.4.2(iv). The main idea of the proof consists in bounding $|\underset{\sim}{W}_{\mu}^{(n)} - W_{\mu}|$. Let $\boldsymbol{Y}_{ki}^{(n)}$ and $\boldsymbol{Y}_{ki}$ stand for $\mathbf{G}_{k,\pm}^{(n)}(\boldsymbol{X}_{ki})$ and $\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})$, respectively. Notice that

$$\begin{aligned}
\underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} &= (n)_m^{-1} \sum_{[i_1,\ldots,i_m]\in I_m^n} k_{f_1,f_2,H}\Big((\boldsymbol{Y}_{1i_1}^{(n)}, \boldsymbol{Y}_{2i_1}^{(n)}), \ldots, (\boldsymbol{Y}_{1i_m}^{(n)}, \boldsymbol{Y}_{2i_m}^{(n)})\Big), \\
W_{J_1,J_2,\mu_{f_1,f_2,H}} &= (n)_m^{-1} \sum_{[i_1,\ldots,i_m]\in I_m^n} k_{f_1,f_2,H}\Big((\boldsymbol{Y}_{1i_1}, \boldsymbol{Y}_{2i_1}), \ldots, (\boldsymbol{Y}_{1i_m}, \boldsymbol{Y}_{2i_m})\Big),
\end{aligned}$$

where

$$\begin{aligned}
&k_{f_1,f_2,H}\Big((\boldsymbol{x}_{11}, \boldsymbol{x}_{21}), \ldots, (\boldsymbol{x}_{1m}, \boldsymbol{x}_{2m})\Big) \\
&\quad := \Big\{ \sum_{\sigma\in H} \mathrm{sgn}(\sigma) f_1(\boldsymbol{x}_{1\sigma(1)}, \ldots, \boldsymbol{x}_{1\sigma(m)})\Big\}\Big\{ \sum_{\sigma\in H} \mathrm{sgn}(\sigma) f_2(\boldsymbol{x}_{2\sigma(1)}, \ldots, \boldsymbol{x}_{2\sigma(m)})\Big\}.
\end{aligned}$$

Since $f_k([\boldsymbol{Y}_{ki_\ell}^{(n)}]_{\ell=1}^m)$ and $f_k([\boldsymbol{Y}_{ki_\ell}]_{\ell=1}^m)$ are almost surely bounded by some constant C_{J_k,f_k}, we deduce

$$\begin{aligned}
&\Big| k_{f_1,f_2,H}\Big([(\boldsymbol{Y}_{1i_\ell}^{(n)}, \boldsymbol{Y}_{2i_\ell}^{(n)})]_{\ell=1}^m\Big) - k_{f_1,f_2,H}\Big([(\boldsymbol{Y}_{1i_\ell}, \boldsymbol{Y}_{2i_\ell})]_{\ell=1}^m\Big)\Big| \\
&\quad \le \mathrm{card}(H)\cdot C_{J_1,f_1}\cdot \sum_{\sigma\in H}\Big| f_2\Big([\boldsymbol{Y}_{2\sigma(i_\ell)}^{(n)}]_{\ell=1}^m\Big) - f_2\Big([\boldsymbol{Y}_{2\sigma(i_\ell)}]_{\ell=1}^m\Big)\Big| \\
&\qquad + \mathrm{card}(H)\cdot C_{J_2,f_2}\cdot \sum_{\sigma\in H}\Big| f_2\Big([\boldsymbol{Y}_{1\sigma(i_\ell)}^{(n)}]_{\ell=1}^m\Big) - f_2\Big([\boldsymbol{Y}_{1\sigma(i_\ell)}]_{\ell=1}^m\Big)\Big|,
\end{aligned}$$

recalling that $\mathrm{card}(H)$ denotes the number of permutations in the subgroup H. Moreover,

$$\begin{aligned}
&\Big| \underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} - W_{J_1,J_2,\mu_{f_1,f_2,H}}\Big| \\
&\quad \le \mathrm{card}(H)^2\cdot C_{J_1,f_1}\cdot \Big[(n)_m^{-1} \sum_{[i_1,\ldots,i_m]\in I_m^n} \Big| f_2\Big([\boldsymbol{Y}_{2i_\ell}^{(n)}]_{\ell=1}^m\Big) - f_2\Big([\boldsymbol{Y}_{2i_\ell}]_{\ell=1}^m\Big)\Big|\Big] \\
&\qquad + \mathrm{card}(H)^2\cdot C_{J_2,f_2}\cdot \Big[(n)_m^{-1} \sum_{[i_1,\ldots,i_m]\in I_m^n} \Big| f_1\Big([\boldsymbol{Y}_{1i_\ell}^{(n)}]_{\ell=1}^m\Big) - f_1\Big([\boldsymbol{Y}_{1i_\ell}]_{\ell=1}^m\Big)\Big|\Big] \xrightarrow{\text{a.s.}} 0.
\end{aligned}$$

This, together with the fact that $W_{J_1,J_2,\mu_{f_1,f_2,H}} \xrightarrow{\text{a.s.}} \mu_\pm(\boldsymbol{X}_1, \boldsymbol{X}_2)$ by the strong consistency of U-statistics, yields $\underset{\sim}{W}^{(n)}_{J_1,J_2,\mu_{f_1,f_2,H}} \xrightarrow{\text{a.s.}} \mu_\pm(\boldsymbol{X}_1, \boldsymbol{X}_2)$. □

B.1.2.3 Proof of Theorem 3.4.1

We first fix some notation and prove a property that will hold for all GSCs μ and associated kernel functions considered in Example 3.2.1(a)–(e). For $k = 1, 2$, let $\boldsymbol{y}^{(n)}_{ki} = \mathbf{J}(\boldsymbol{u}^{*(n)}_{ki})$, where $\boldsymbol{u}^{*(n)}_{ki}$, $i \in [\![n]\!]$ are the deterministic points forming the grid $\mathfrak{G}^{d_k}_{\boldsymbol{n}}$. Writing $\boldsymbol{Y}^{(n)}_{ki}$ and $\boldsymbol{Y}_{ki}$ for $\mathbf{G}^{(n)}_{k,\pm}(\boldsymbol{X}_{ki})$ and $\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})$, respectively, let us show that

$$\Xi^{(n)}_k := \sup_{1\le i\le n} \|\boldsymbol{Y}^{(n)}_{ki} - \boldsymbol{Y}_{ki}\| \xrightarrow{\text{a.s.}} 0, \quad k = 1, 2. \tag{B.1.1}$$

Recall that, by definition of strong regularity, J_k is Lipschitz-continuous with some constant L_k, strictly monotone, and satisfies $J_k(0) = 0$. Then we immediately have $|J_k(u)| \le L_k$ for all $u \in [0, 1)$, and thus $\boldsymbol{Y}^{(n)}_{ki}$ and $\boldsymbol{Y}_{ki}$ are almost surely bounded by L_k. As long as $\mathrm{P}_{\boldsymbol{X}_k} \in \mathcal{P}^{\#}_{d_k}$, in order to prove that $\Xi^{(n)}_k \xrightarrow{\text{a.s.}} 0$, it suffices to show that $\|\mathbf{J}_k(\boldsymbol{u}_{k1}) - \mathbf{J}_k(\boldsymbol{u}_{k2})\| \le 2L_k\|\boldsymbol{u}_{k1} - \boldsymbol{u}_{k2}\|$ for any $\boldsymbol{u}_{k1}, \boldsymbol{u}_{k2} \in \mathbb{R}^{d_k}$ with $\|\boldsymbol{u}_{k1}\|, \|\boldsymbol{u}_{k2}\| < 1$. Without loss of generality, assume that $\|\boldsymbol{u}_{k2}\| \le \|\boldsymbol{u}_{k1}\|$. If $\|\boldsymbol{u}_{k2}\| = 0$, the claim is obvious by noticing $|J_k(u)| \le L_k u$ for $u \in [0, 1)$ and then $\|\mathbf{J}_k(\boldsymbol{u}_{k1})\| \le L_k\|\boldsymbol{u}_{k1}\|$; otherwise we have

$$\begin{aligned}
\|\mathbf{J}_k(\boldsymbol{u}_{k1}) - \mathbf{J}_k(\boldsymbol{u}_{k2})\| &\le \Big\|\mathbf{J}_k(\boldsymbol{u}_{k1}) - \mathbf{J}_k\Big(\frac{\|\boldsymbol{u}_{k2}\|}{\|\boldsymbol{u}_{k1}\|}\boldsymbol{u}_{k1}\Big)\Big\| + \Big\|\mathbf{J}_k\Big(\frac{\|\boldsymbol{u}_{k2}\|}{\|\boldsymbol{u}_{k1}\|}\boldsymbol{u}_{k1}\Big) - \mathbf{J}_k(\boldsymbol{u}_{k2})\Big\| \\
&= \Big|J_k(\|\boldsymbol{u}_{k1}\|) - J_k(\|\boldsymbol{u}_{k2}\|)\Big| + \frac{J_k(\|\boldsymbol{u}_{k2}\|)}{\|\boldsymbol{u}_{k2}\|} \cdot \Big\|\frac{\|\boldsymbol{u}_{k2}\|}{\|\boldsymbol{u}_{k1}\|}\boldsymbol{u}_{k1} - \boldsymbol{u}_{k2}\Big\| \\
&\le L_k\Big|\|\boldsymbol{u}_{k1}\| - \|\boldsymbol{u}_{k2}\|\Big| + L_k\Big\|\frac{\|\boldsymbol{u}_{k2}\|}{\|\boldsymbol{u}_{k1}\|}\boldsymbol{u}_{k1} - \boldsymbol{u}_{k2}\Big\| \le 2L_k\|\boldsymbol{u}_{k1} - \boldsymbol{u}_{k2}\|.
\end{aligned}$$

This completes the proof of (B.1.1).

B.1.2.3.1 Proof of Theorem 3.4.1 ($h = h_{\mathrm{dCov}^2}$)

Proof of Theorem 3.4.1 ($h = h_{\mathrm{dCov}^2}$). Recall that

$$f_1^{\mathrm{dCov}}([\boldsymbol{w}_i]_{i=1}^4) = \frac{1}{2}\|\boldsymbol{w}_1 - \boldsymbol{w}_2\| \quad \text{and} \quad f_2^{\mathrm{dCov}}([\boldsymbol{w}_i]_{i=1}^4) = \frac{1}{2}\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|,$$

with possibly different dimension for the inputs. Now, $\frac{1}{2}\|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_2}^{(n)}\|$ and $\frac{1}{2}\|\boldsymbol{Y}_{ki_1} - \boldsymbol{Y}_{ki_2}\|$ are almost surely bounded by L_k, since $\boldsymbol{Y}_{ki}^{(n)}$ and $\boldsymbol{Y}_{ki}$ are. Next,

$$\left|\frac{1}{2}\|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_2}^{(n)}\| - \frac{1}{2}\|\boldsymbol{Y}_{ki_1} - \boldsymbol{Y}_{ki_2}\|\right| \le \frac{1}{2}\|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_1}\| + \frac{1}{2}\|\boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_2}\| \le \sup_{1\le i\le n}\|\boldsymbol{Y}_{ki}^{(n)} - \boldsymbol{Y}_{ki}\|,$$

and we deduce that

$$(n)_4^{-1}\sum_{[i_1,\ldots,i_4]\in I_4^n}\left|\frac{1}{2}\|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_2}^{(n)}\| - \frac{1}{2}\|\boldsymbol{Y}_{ki_1} - \boldsymbol{Y}_{ki_2}\|\right| \le \sup_{1\le i\le n}\|\boldsymbol{Y}_{ki}^{(n)} - \boldsymbol{Y}_{ki}\| \xrightarrow{\text{a.s.}} 0.$$

Both conditions in (3.4.4) are satisfied, and the proof is thus completed. □

B.1.2.3.2 Proof of Theorem 3.4.1 ($h = h_M$)

Proof of Theorem 3.4.1 ($h = h_M$). Recall that $f_1^M([\boldsymbol{w}_i]_{i=1}^5) = f_2^M([\boldsymbol{w}_i]_{i=1}^5) = \frac{1}{2}\mathbb{1}(\boldsymbol{w}_1, \boldsymbol{w}_2 \preceq \boldsymbol{w}_5)$, up to a change in input dimension for the two functions. It is obvious that $f_k(\{\boldsymbol{Y}_{ki_\ell}^{(n)}\}_{\ell=1}^m)$ and $f_k(\{\boldsymbol{Y}_{ki_\ell}\}_{\ell=1}^m)$ are almost surely bounded. Next we verify the second condition in (3.4.4).

We have for $k = 1, 2$,

$$\left|\mathbb{1}(\boldsymbol{Y}_{ki_1}^{(n)}, \boldsymbol{Y}_{ki_2}^{(n)} \preceq \boldsymbol{Y}_{ki_5}^{(n)}) - \mathbb{1}(\boldsymbol{Y}_{ki_1}, \boldsymbol{Y}_{ki_2} \preceq \boldsymbol{Y}_{ki_5})\right| \le \mathbb{1}(\mathcal{B}_{k;i_1,i_2,i_3,i_4,i_5}^{\complement}),$$

where

$$\mathcal{B}_{k;i_1,i_2,i_3,i_4,i_5} := \left\{\|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| \ge 2\,\Xi_k^{(n)}, \quad \|\boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| \ge 2\,\Xi_k^{(n)}\right\}.$$

Accordingly,

$$\begin{aligned}
&(n)_5^{-1}\sum_{[i_1,\ldots,i_5]\in I_5^n}\left|\mathbb{1}(\boldsymbol{Y}_{ki_1}^{(n)}, \boldsymbol{Y}_{ki_2}^{(n)} \preceq \boldsymbol{Y}_{ki_5}^{(n)}) - \mathbb{1}(\boldsymbol{Y}_{ki_1}, \boldsymbol{Y}_{ki_2} \preceq \boldsymbol{Y}_{ki_5})\right| \\
&\quad \le (n)_3^{-1}\mathrm{card}\left\{[i_1, i_2, i_5] \in I_3^n : \|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| < 2\,\Xi_k^{(n)} \text{ or } \quad \|\boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| < 2\,\Xi_k^{(n)}\right\}
\end{aligned}$$

$$= (n)_3^{-1}\mathrm{card}\Big\{[i_1, i_2, i_5] \in I_3^n : \|\boldsymbol{y}_{ki_1}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < 2\,\Xi_k^{(n)} \quad \text{or} \quad \|\boldsymbol{y}_{ki_2}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < 2\,\Xi_k^{(n)}\Big\} \xrightarrow{\text{a.s.}} 0, \tag{B.1.2}$$

which completes the proof. □

B.1.2.3.3 Proof of Theorem 3.4.1 ($h = h_D$)

Proof of Theorem 3.4.1 ($h = h_D$). Recall that $f_1^D([\boldsymbol{w}_i]_{i=1}^5) = f_2^D([\boldsymbol{w}_i]_{i=1}^5) = \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_5, \boldsymbol{w}_2 - \boldsymbol{w}_5)$, up to a change in input dimension for the two functions. Obviously, $f_k([\boldsymbol{Y}_{ki_\ell}^{(n)}]_{\ell=1}^m)$ and $f_k([\boldsymbol{Y}_{ki_\ell}]_{\ell=1}^m)$ are almost surely bounded. To verify the second condition in (3.4.4), we start by bounding the difference between $\mathsf{Arc}(\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}, \boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)})$ and $\mathsf{Arc}(\boldsymbol{Y}_{ki_1} - \boldsymbol{Y}_{ki_5}, \boldsymbol{Y}_{ki_2} - \boldsymbol{Y}_{ki_5})$.

For $k = 1, 2$, consider $(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}, \boldsymbol{y}_{k5}) \in (\mathbb{R}^{d_k})^3$ such that

$$\min\{\|\boldsymbol{y}_{k1} - \boldsymbol{y}_{k5}\|, \|\boldsymbol{y}_{k2} - \boldsymbol{y}_{k5}\|\} \geq \eta \quad \text{and} \quad \zeta \leq \mathsf{Arc}(\boldsymbol{y}_{k1} - \boldsymbol{y}_{k5}, \boldsymbol{y}_{k2} - \boldsymbol{y}_{k5}) \leq \frac{1}{2} - \zeta,$$

where η and ζ will be specified later on. For $(\boldsymbol{y}'_{k1}, \boldsymbol{y}'_{k2}, \boldsymbol{y}'_{k5}) \in (\mathbb{R}^{d_k})^3$ satisfying $\|\boldsymbol{y}_{ki} - \boldsymbol{y}'_{ki}\| \leq \delta$ for $i = 1, 2, 5$,

$$\begin{aligned}
&\mathsf{Arc}(\boldsymbol{y}_{k1} - \boldsymbol{y}_{k5}, \boldsymbol{y}_{k1} - \boldsymbol{y}'_{k5}) \leq \frac{1}{2\pi}\arcsin\frac{\|\boldsymbol{y}_{k5} - \boldsymbol{y}'_{k5}\|}{\|\boldsymbol{y}_{k1} - \boldsymbol{y}_{k5}\|} \leq \frac{1}{2\pi}\arcsin\frac{\delta}{\eta},\\
&\mathsf{Arc}(\boldsymbol{y}_{k1} - \boldsymbol{y}'_{k5}, \boldsymbol{y}'_{k1} - \boldsymbol{y}'_{k5}) \leq \frac{1}{2\pi}\arcsin\frac{\|\boldsymbol{y}_{k1} - \boldsymbol{y}'_{k1}\|}{\|\boldsymbol{y}_{k1} - \boldsymbol{y}'_{k5}\|} \leq \frac{1}{2\pi}\arcsin\frac{\delta}{\eta - \delta},\\
&\mathsf{Arc}(\boldsymbol{y}_{k2} - \boldsymbol{y}_{k5}, \boldsymbol{y}_{k2} - \boldsymbol{y}'_{k5}) \leq \frac{1}{2\pi}\arcsin\frac{\|\boldsymbol{y}_{k5} - \boldsymbol{y}'_{k5}\|}{\|\boldsymbol{y}_{k2} - \boldsymbol{y}_{k5}\|} \leq \frac{1}{2\pi}\arcsin\frac{\delta}{\eta},\\
\text{and}\quad &\mathsf{Arc}(\boldsymbol{y}_{k2} - \boldsymbol{y}'_{k5}, \boldsymbol{y}'_{k2} - \boldsymbol{y}'_{k5}) \leq \frac{1}{2\pi}\arcsin\frac{\|\boldsymbol{y}_{k2} - \boldsymbol{y}'_{k2}\|}{\|\boldsymbol{y}_{k2} - \boldsymbol{y}'_{k5}\|} \leq \frac{1}{2\pi}\arcsin\frac{\delta}{\eta - \delta}.
\end{aligned}$$

Assuming that

$$\frac{1}{2\pi}\Big(2\arcsin\frac{\delta}{\eta} + 2\arcsin\frac{\delta}{\eta - \delta}\Big) \leq \zeta, \tag{B.1.3}$$

we obtain

$$|\mathsf{Arc}(\boldsymbol{y}_{k1}-\boldsymbol{y}_{k5},\boldsymbol{y}_{k2}-\boldsymbol{y}_{k5})-\mathsf{Arc}(\boldsymbol{y}'_{k1}-\boldsymbol{y}'_{k5},\boldsymbol{y}'_{k2}-\boldsymbol{y}'_{k5})|\le\frac{1}{2\pi}\Big(2\arcsin\frac{\delta}{\eta}+2\arcsin\frac{\delta}{\eta-\delta}\Big).$$

For $\delta\le 1/4$, take $\eta=\sqrt{\delta}$ and $\zeta=3\sqrt{\delta}/2$ such that (B.1.3) holds,

$$\begin{aligned}\frac{1}{2\pi}\Big(2\arcsin\frac{\delta}{\eta}+2\arcsin\frac{\delta}{\eta-\delta}\Big)&=\frac{1}{2\pi}\Big(2\arcsin\sqrt{\delta}+2\arcsin\frac{\sqrt{\delta}}{1-\sqrt{\delta}}\Big)\\ \le\frac{1}{2\pi}\Big(2\arcsin\sqrt{\delta}+2\arcsin 2\sqrt{\delta}\Big)&\le\frac{1}{2\pi}\Big(2\frac{\pi}{2}\sqrt{\delta}+2\frac{\pi}{2}(2\sqrt{\delta})\Big)=\frac{3}{2}\sqrt{\delta}=\zeta.\end{aligned}$$

It follows that for $\delta\le 1/4$ and $(\boldsymbol{y}_{k1},\boldsymbol{y}_{k2},\boldsymbol{y}_{k5}),(\boldsymbol{y}'_{k1},\boldsymbol{y}'_{k2},\boldsymbol{y}'_{k5})\in(\mathbb{R}^{d_k})^3$ such that

$$\begin{aligned}&\min\{\|\boldsymbol{y}_{k1}-\boldsymbol{y}_{k5}\|,\|\boldsymbol{y}_{k2}-\boldsymbol{y}_{k5}\|\}\ge\sqrt{\delta},\quad \frac{3}{2}\sqrt{\delta}\le\mathsf{Arc}(\boldsymbol{y}_{k1}-\boldsymbol{y}_{k5},\boldsymbol{y}_{k2}-\boldsymbol{y}_{k5})\le\frac{1}{2}-\frac{3}{2}\sqrt{\delta},\\ &\qquad\text{and}\qquad \|\boldsymbol{y}_{ki}-\boldsymbol{y}'_{ki}\|\le\delta\qquad\text{for } i=1,2,5,\end{aligned}$$

we have

$$|\mathsf{Arc}(\boldsymbol{y}_{k1}-\boldsymbol{y}_{k5},\boldsymbol{y}_{k2}-\boldsymbol{y}_{k5})-\mathsf{Arc}(\boldsymbol{y}'_{k1}-\boldsymbol{y}'_{k5},\boldsymbol{y}'_{k2}-\boldsymbol{y}'_{k5})|\le\frac{3}{2}\sqrt{\delta}.$$

Then, for $k=1,2$,

$$\begin{aligned}&\Big|\mathsf{Arc}(\boldsymbol{Y}^{(n)}_{ki_1}-\boldsymbol{Y}^{(n)}_{ki_5},\boldsymbol{Y}^{(n)}_{ki_2}-\boldsymbol{Y}^{(n)}_{ki_5})-\mathsf{Arc}(\boldsymbol{Y}_{ki_1}-\boldsymbol{Y}_{ki_5},\boldsymbol{Y}_{ki_2}-\boldsymbol{Y}_{ki_5})\Big|\\ &\le\frac{3}{2}\sqrt{\Xi^{(n)}_k}\cdot\mathbb{1}(\mathcal{A}_{k;i_1,i_2,i_3,i_4,i_5})+\Big(\frac{1}{2}+\frac{1}{2}\Big)\cdot\mathbb{1}(\mathcal{A}^{\complement}_{k;i_1,i_2,i_3,i_4,i_5})\le\frac{3}{2}\sqrt{\Xi^{(n)}_k}+\mathbb{1}(\mathcal{A}^{\complement}_{k;i_1,i_2,i_3,i_4,i_5}),\end{aligned}$$

where

$$\begin{aligned}\mathcal{A}_{k;i_1,i_2,i_3,i_4,i_5}:=\Big\{&\Xi^{(n)}_k\le\frac{1}{4},\quad\|\boldsymbol{Y}^{(n)}_{ki_1}-\boldsymbol{Y}^{(n)}_{ki_5}\|\ge\sqrt{\Xi^{(n)}_k},\quad\|\boldsymbol{Y}^{(n)}_{ki_2}-\boldsymbol{Y}^{(n)}_{ki_5}\|\ge\sqrt{\Xi^{(n)}_k},\\ &\text{and}\quad\frac{3}{2}\sqrt{\Xi^{(n)}_k}\le\mathsf{Arc}(\boldsymbol{Y}^{(n)}_{ki_1}-\boldsymbol{Y}^{(n)}_{ki_5},\boldsymbol{Y}^{(n)}_{ki_2}-\boldsymbol{Y}^{(n)}_{ki_5})\le\frac{1}{2}-\frac{3}{2}\sqrt{\Xi^{(n)}_k}\Big\},\end{aligned}$$

and, accordingly,

$$
\begin{aligned}
&(n)_5^{-1} \sum_{[i_1,\dots,i_5]\in I_5^n} \Big| \frac{1}{2}\mathsf{Arc}(\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}, \boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}) - \frac{1}{2}\mathsf{Arc}(\boldsymbol{Y}_{ki_1} - \boldsymbol{Y}_{ki_5}, \boldsymbol{Y}_{ki_2} - \boldsymbol{Y}_{ki_5}) \Big| \\
&\quad \le \frac{1}{2}\Big(\frac{3}{2}\sqrt{\Xi_k^{(n)}} + \mathbb{1}\Big\{\Xi_k^{(n)} > \frac{1}{4}\Big\} \\
&\qquad + (n)_3^{-1}\text{card}\Big\{[i_1, i_2, i_5] \in I_3^n : \|\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| < \sqrt{\Xi_k^{(n)}},\ \text{or } \|\boldsymbol{Y}_{ki_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}\| < \sqrt{\Xi_k^{(n)}}, \\
&\qquad\qquad \text{or } \mathsf{Arc}(\boldsymbol{Y}_{ki_1}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}, \boldsymbol{Y}_{i_2}^{(n)} - \boldsymbol{Y}_{ki_5}^{(n)}) \in \Big[0, \frac{3}{2}\sqrt{\Xi_k^{(n)}}\Big) \cup \Big(\frac{1}{2} - \frac{3}{2}\sqrt{\Xi_k^{(n)}}, \frac{1}{2}\Big]\Big\}\Big) \\
&\quad = \frac{1}{2}\Big(\frac{3}{2}\sqrt{\Xi_k^{(n)}} + \mathbb{1}\Big\{\Xi_k^{(n)} > \frac{1}{4}\Big\} \\
&\qquad + (n)_3^{-1}\text{card}\Big\{[i_1, i_2, i_5] \in I_3^n : \|\boldsymbol{y}_{ki_1}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < \sqrt{\Xi_k^{(n)}},\ \text{or } \|\boldsymbol{y}_{ki_2}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < \sqrt{\Xi_k^{(n)}}, \\
&\qquad\qquad \text{or } \mathsf{Arc}(\boldsymbol{y}_{ki_1}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}, \boldsymbol{y}_{ki_2}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}) \in \Big[0, \frac{3}{2}\sqrt{\Xi_k^{(n)}}\Big) \cup \Big(\frac{1}{2} - \frac{3}{2}\sqrt{\Xi_k^{(n)}}, \frac{1}{2}\Big]\Big\}\Big).
\end{aligned}
\tag{B.1.4}
$$

Since, for any sequence $[\delta^{(n)}]_{n=1}^{\infty}$ tending to 0, it holds that

$$
\begin{aligned}
&(n)_3^{-1}\text{card}\Big\{[i_1, i_2, i_5] \in I_3^n : \|\boldsymbol{y}_{ki_1}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < \sqrt{\delta^{(n)}}, \quad \text{or} \quad \|\boldsymbol{y}_{ki_2}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}\| < \sqrt{\delta^{(n)}}, \\
&\qquad \text{or} \quad \mathsf{Arc}(\boldsymbol{y}_{ki_1}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}, \boldsymbol{y}_{ki_2}^{(n)} - \boldsymbol{y}_{ki_5}^{(n)}) \in \Big[0, \frac{3}{2}\sqrt{\delta^{(n)}}\Big) \cup \Big(\frac{1}{2} - \frac{3}{2}\sqrt{\delta^{(n)}}, \frac{1}{2}\Big]\Big\} \to 0,
\end{aligned}
$$

we have shown that (B.1.4) converges to 0 almost surely. This completes the proof. □

B.1.2.3.4 Proof of Theorem 3.4.1 ($h = h_R$, h_{τ^*})

Proof of Theorem 3.4.1 ($h = h_R$, h_{τ^*}). The proof is similar to the proof of Theorem 3.4.1 ($h = h_D$) and hence omitted. □

B.1.3 Proofs for Section 3.5

B.1.3.1 Proof of Proposition 3.5.1

Proof of Proposition 3.5.1. In view of Lemma 3 in Weihs et al. (2018), the claim readily follows from the theory of degenerate U-statistics (Serfling, 1980, Chap. 5.5.2). □

B.1.3.2 Proof of Theorem 3.5.1

Proof of Theorem 3.5.1. For $k = 1, 2$, let $\mathrm{P}^{(n)}_{J_k,d_k}$ and P_{J_k,d_k} denote the distributions of $\boldsymbol{W}^{(n)}_{k1}$ and $\boldsymbol{W}_{k1}$, respectively, and let again $\boldsymbol{Y}^{(n)}_{ki}$ and $\boldsymbol{Y}_{ki}$ stand for $\mathbf{G}^{(n)}_{k,\pm}(\boldsymbol{X}_{ki})$ and $\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})$, respectively. Consider the Hoeffding decomposition

$$\underset{\sim}{W}^{(n)}_{\mu} = \sum_{\ell=1}^{m} \underbrace{\binom{m}{\ell}\binom{n}{\ell}^{-1} \sum_{1\le i_1<\cdots<i_\ell\le n} \widetilde{h}_{\mu,\ell}\Big((\mathbf{Y}^{(n)}_{1i_1}, \mathbf{Y}^{(n)}_{2i_1}), \dots, (\mathbf{Y}^{(n)}_{1i_\ell}, \mathbf{Y}^{(n)}_{2i_\ell}); \mathrm{P}^{(n)}_{J_1,d_1} \otimes \mathrm{P}^{(n)}_{J_2,d_2}\Big)}_{\underset{\sim}{H}_{n,\ell}}, \tag{B.1.5}$$

of $\underset{\sim}{W}^{(n)}_{\mu}$ with respect to the product measure $\mathrm{P}^{(n)}_{J_1,d_1} \otimes \mathrm{P}^{(n)}_{J_2,d_2}$ and the Hoeffding decomposition

$$W_{\mu} = \sum_{\ell=1}^{m} \underbrace{\binom{m}{\ell}\binom{n}{\ell}^{-1} \sum_{1\le i_1<\cdots<i_\ell\le n} \widetilde{h}_{\mu,\ell}\Big((\mathbf{Y}_{1i_1}, \mathbf{Y}_{2i_1}), \dots, (\mathbf{Y}_{1i_\ell}, \mathbf{Y}_{2i_\ell}); \mathrm{P}_{J_1,d_1} \otimes \mathrm{P}_{J_2,d_2}\Big)}_{H_{n,\ell}}. \tag{B.1.6}$$

of W_μ with respect to product measure $\mathrm{P}_{J_1,d_1} \otimes \mathrm{P}_{J_2,d_2}$.

The proof is divided into three steps. The first step shows that $n\underset{\sim}{H}_{n,1} = nH_{n,1} = 0$, the second step that $n\underset{\sim}{H}_{n,2} - nH_{n,2} = o_{\mathrm{P}}(1)$. The third step verifies that $n\underset{\sim}{H}_{n,\ell}$ and $nH_{n,\ell}$, $\ell = 3, 4, \dots, m$ all are $o_{\mathrm{P}}(1)$ terms.

Step I. Lemma 3 in Weihs et al. (2018) confirms that

$$\widetilde{h}_{\mu,1}(\cdot; \mathrm{P}^{(n)}_{J_1,d_1} \otimes \mathrm{P}^{(n)}_{J_2,d_2}) = 0 = \widetilde{h}_{\mu,1}(\cdot; \mathrm{P}_{J_1,d_1} \otimes \mathrm{P}_{J_2,d_2}),$$

and thus $n\underset{\sim}{H}_{n,1} = nH_{n,1} = 0$.

Step II. Lemma 3 in Weihs et al. (2018) shows that,

$$\binom{m}{2} \cdot \widetilde{h}_{\mu,2}\Big((\boldsymbol{y}_{11}, \boldsymbol{y}_{21}), (\boldsymbol{y}_{12}, \boldsymbol{y}_{22}); \mathrm{P}^{(n)}_{J_1,d_1} \otimes \mathrm{P}^{(n)}_{J_2,d_2}\Big) = g_1^{(n)}(\boldsymbol{y}_{11}, \boldsymbol{y}_{12}) g_2^{(n)}(\boldsymbol{y}_{21}, \boldsymbol{y}_{22}),$$

$$\text{and} \quad \binom{m}{2} \cdot \widetilde{h}_{\mu,2}\Big((\boldsymbol{y}_{11}, \boldsymbol{y}_{21}), (\boldsymbol{y}_{12}, \boldsymbol{y}_{22}); \mathrm{P}_{J_1,d_1} \otimes \mathrm{P}_{J_2,d_2}\Big) = g_1(\boldsymbol{y}_{11}, \boldsymbol{y}_{12}) g_2(\boldsymbol{y}_{21}, \boldsymbol{y}_{22}),$$

where $g_k^{(n)}$ and g_k are defined in (3.5.5) and (3.5.2). To prove that $n\underset{\sim}{H}_{n,2} - nH_{n,2} = o_{\mathrm{P}}(1)$, it suffices to show that

$$\begin{aligned} &\mathrm{E}\big[(n\underset{\sim}{H}_{n,2} - nH_{n,2})^2\big] \\ &= \mathrm{E}\Big[\Big(\frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1^{(n)}(\boldsymbol{Y}_{1i}^{(n)}, \boldsymbol{Y}_{1j}^{(n)}) g_2^{(n)}(\boldsymbol{Y}_{2i}^{(n)}, \boldsymbol{Y}_{2j}^{(n)}) - \frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{1j}) g_2(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{2j})\Big)^2\Big] \\ &= o(1). \end{aligned} \tag{B.1.7}$$

We proceed in three sub-steps.

Step II-1. The theory of degenerate U-statistics (cf. Equation (7) of Section 1.6 in Lee (1990)) yields that

$$\mathrm{E}\big[(nH_{n,2})^2\big] = \frac{2n}{n-1}\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12})^2\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22})^2\big]. \tag{B.1.8}$$

Step II-2. We next deduce that

$$\begin{aligned} &\mathrm{E}\big[(n\underset{\sim}{H}_{n,2})(nH_{n,2})\big] \\ &= \mathrm{E}\Big[\Big(\frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1^{(n)}(\boldsymbol{Y}_{1i}^{(n)}, \boldsymbol{Y}_{1j}^{(n)}) g_2^{(n)}(\boldsymbol{Y}_{2i}^{(n)}, \boldsymbol{Y}_{2j}^{(n)})\Big)\Big(\frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{1j}) g_2(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{2j})\Big)\Big] \\ &\to 2\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12})^2\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22})^2\big]. \end{aligned} \tag{B.1.9}$$

By symmetry, we have

$$\begin{aligned}
\mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{ki}^{(n)}, \boldsymbol{Y}_{kj}^{(n)})g_k(\boldsymbol{Y}_{ki}, \boldsymbol{Y}_{kj})\big] &= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] =: A_k^{(n)}, \\
\mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k\ell}^{(n)}, \boldsymbol{Y}_{kj}^{(n)})g_k(\boldsymbol{Y}_{ki}, \boldsymbol{Y}_{kj})\big] &= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{ki}^{(n)}, \boldsymbol{Y}_{kr}^{(n)})g_k(\boldsymbol{Y}_{ki}, \boldsymbol{Y}_{kj})\big] \\
&= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k3}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] =: B_k^{(n)}, \\
\mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k\ell}^{(n)}, \boldsymbol{Y}_{kr}^{(n)})g_k(\boldsymbol{Y}_{ki}, \boldsymbol{Y}_{kj})\big] &= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k3}^{(n)}, \boldsymbol{Y}_{k4}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] =: C_k^{(n)}
\end{aligned}$$

for all distinct i, j, ℓ, r, and also

$$A_k^{(n)} = \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big], \tag{B.1.10}$$

$$\begin{aligned}
A_k^{(n)} + (n-2)B_k^{(n)} &= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] + \sum_{\ell:\ell\neq 1,2} \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k\ell}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k2}, \boldsymbol{Y}_{k2})\big] \\
&= -\,\mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k2}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big],
\end{aligned} \tag{B.1.11}$$

$$\begin{aligned}
2B_k^{(n)} + (n-3)C_k^{(n)} &= \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k3}^{(n)}, \boldsymbol{Y}_{k1}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] + \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k3}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] \\
&\quad + \sum_{\ell:\ell\neq 1,2,3} \mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k3}^{(n)}, \boldsymbol{Y}_{k\ell}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] \\
&= -\,\mathrm{E}\big[g_k^{(n)}(\boldsymbol{Y}_{k3}^{(n)}, \boldsymbol{Y}_{k3}^{(n)})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big].
\end{aligned} \tag{B.1.12}$$

We claim that

$$A_k^{(n)} \to \mathrm{E}\big[g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})^2\big], \tag{B.1.13}$$

$$A_k^{(n)} + (n-2)B_k^{(n)} \to -\,\mathrm{E}\big[g_k(\boldsymbol{Y}_{k2}, \boldsymbol{Y}_{k2})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] = 0, \tag{B.1.14}$$

$$2B_k^{(n)} + (n-3)C_k^{(n)} \to -\,\mathrm{E}\big[g_k(\boldsymbol{Y}_{k3}, \boldsymbol{Y}_{k3})g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\big] = 0. \tag{B.1.15}$$

We only prove (B.1.13), as (B.1.14) and (B.1.15) are quite similar.

If Condition (3.5.6) holds, we obtain, since $\mathrm{E}\big[f_k([\boldsymbol{W}_{ki_\ell}]_{\ell=1}^m)^2\big] < \infty$, that

$$\|g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\|_{\mathsf{L}^1} \le \|g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\|_{\mathsf{L}^2} < \infty.$$

To prove (B.1.13), we still need to show that $\boldsymbol{Y}_{ki}^{(n)} \xrightarrow{\mathsf{L}^2} \boldsymbol{Y}_{ki}$ for $k = 1, 2$. Since the scores J_k, $k = 1, 2$ are weakly regular (cf. Definition 3.4.2) and square-integrable, we obtain

$$\lim_{n\to\infty} n^{-1} \sum_{r=1}^{n} J^2\Big(\frac{r}{n+1}\Big) = \int_0^1 J^2(u)\mathrm{d}u,$$

and thus $\mathrm{E}\|\boldsymbol{Y}_{ki}^{(n)}\|^2 \to \mathrm{E}\|\boldsymbol{Y}_{ki}\|^2$. Notice also that $\boldsymbol{Y}_{ki}^{(n)} \xrightarrow{\mathsf{a.s.}} \boldsymbol{Y}_{ki}$. Using Vitali's theorem (Shorack, 2017, Chap. 3, Theorem 5.5) yields $\mathrm{E}\|\boldsymbol{Y}_{ki}^{(n)} - \boldsymbol{Y}_{ki}\|^2 \to 0$.

Because $g_k^{(n)}(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) \rightrightarrows g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2})$, we have

$$\begin{aligned}&\mathrm{E}\big[|g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)}) - g_k(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})| \cdot |g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})|\big]\\ &\le \|g_k^{(n)}(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)}) - g_k(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)})\|_{\mathsf{L}^\infty} \cdot \|g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\|_{\mathsf{L}^1} \to 0. \qquad \text{(B.1.16)}\end{aligned}$$

Next, since g_k is Lipschitz-continuous, by the fact that $\boldsymbol{Y}_{ki}^{(n)} \xrightarrow{\mathsf{L}^2} \boldsymbol{Y}_{ki}$,

$$\begin{aligned}&\mathrm{E}\big[|g_k(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)}) - g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})| \cdot |g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})|\big]\\ &\le \|g_k(\boldsymbol{Y}_{k1}^{(n)}, \boldsymbol{Y}_{k2}^{(n)}) - g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\|_{\mathsf{L}^2} \cdot \|g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})\|_{\mathsf{L}^2} \to 0; \qquad \text{(B.1.17)}\end{aligned}$$

Combining (B.1.16) and (B.1.17) yields (B.1.13).

Having established (B.1.13)–(B.1.15), we obtain that

$$A_k^{(n)} \to \mathrm{E}\big[g_k(\boldsymbol{Y}_{k1}, \boldsymbol{Y}_{k2})^2\big], \quad B_k^{(n)} = O(n^{-1}) \quad \text{and} \quad C_k^{(n)} = o(n^{-1}). \qquad \text{(B.1.18)}$$

Plugging (B.1.18) into the left-hand side of (B.1.9) gives

$$\begin{aligned}&\mathrm{E}\Big[\Big(\frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1^{(n)}(\boldsymbol{Y}_{1i}^{(n)}, \boldsymbol{Y}_{1j}^{(n)}) g_2^{(n)}(\boldsymbol{Y}_{2i}^{(n)}, \boldsymbol{Y}_{2j}^{(n)})\Big)\Big(\frac{1}{n-1}\sum_{(i,j)\in I_2^n} g_1(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{1j}) g_2(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{2j})\Big)\Big]\\ &= \frac{n(n-1)}{(n-1)^2}\Big\{2A_1^{(n)}A_2^{(n)} + 4(n-2)B_1^{(n)}B_2^{(n)} + (n-2)(n-3)C_1^{(n)}C_2^{(n)}\Big\}\\ &\to 2\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12})^2\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22})^2\big].\end{aligned}$$

This completes the proof of (B.1.9).

Step II-3. In order to prove (B.1.7), it remains to show that

$$\mathrm{E}\big[(n\underset{\sim}{H}_{n,2})^2\big] \to 2\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12})^2\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22})^2\big]. \tag{B.1.19}$$

Notice that $n\underset{\sim}{H}_{n,2}$ is a double-indexed permutation statistic. Applying Equations (2.2)–(2.3) in Barbour and Eagleson (1986) yields $\mathrm{E}\big[n\underset{\sim}{H}_{n,2}\big] = n\mu_1^{(n)}\mu_2^{(n)}$, and

$$\begin{aligned}\mathrm{Var}(n\underset{\sim}{H}_{n,2}) = &\frac{4n^2}{(n-1)^3(n-2)^2}\Big(\frac{\sum_{i=1}^n \{\zeta_{1i}^{(n)}\}^2}{n}\Big)\Big(\frac{\sum_{i=1}^n \{\zeta_{2i}^{(n)}\}^2}{n}\Big)\\ &+\frac{2n}{n-3}\Big(\frac{\sum_{i\neq j}\{\eta_{1ij}^{(n)}\}^2}{n(n-1)}\Big)\Big(\frac{\sum_{i\neq j}\{\eta_{2ij}^{(n)}\}^2}{n(n-1)}\Big),\end{aligned}$$

where for $k = 1, 2$,

$$\begin{aligned}\mu_k^{(n)} &:= \frac{1}{n(n-1)}\sum_{i\neq j} g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{kj}^{(n)}),\\ \zeta_{ki}^{(n)} &:= \sum_{j:j\neq i}\Big\{g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{kj}^{(n)}) - \mu_k^{(n)}\Big\},\\ \eta_{kij}^{(n)} &:= g_k^{(n)}(\boldsymbol{y}_i^{(n)}, \boldsymbol{y}_j^{(n)}) - \frac{\zeta_{ki}^{(n)}}{n-2} - \frac{\zeta_{kj}^{(n)}}{n-2} - \mu_k^{(n)}.\end{aligned}$$

Direct computation gives

$$\begin{aligned}\mu_k^{(n)} &= -\frac{1}{n(n-1)}\sum_{i=1}^n g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{ki}^{(n)}),\\ \zeta_{ki}^{(n)} &= -g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{ki}^{(n)}) + \frac{1}{n}\sum_{j=1}^n g_k^{(n)}(\boldsymbol{y}_{kj}^{(n)}, \boldsymbol{y}_{kj}^{(n)}),\\ \eta_{kij}^{(n)} &= g_k^{(n)}(\boldsymbol{y}_i^{(n)}, \boldsymbol{y}_j^{(n)}) + \frac{g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{ki}^{(n)})}{n-2} + \frac{g_k^{(n)}(\boldsymbol{y}_{kj}^{(n)}, \boldsymbol{y}_{kj}^{(n)})}{n-2} - \frac{1}{(n-1)(n-2)}\sum_{i=1}^n g_k^{(n)}(\boldsymbol{y}_{ki}^{(n)}, \boldsymbol{y}_{ki}^{(n)}).\end{aligned}$$

Moreover, we can write $\mathrm{E}\big[n\underset{\sim}{H}_{n,2}\big]$ and $\mathrm{Var}(n\underset{\sim}{H}_{n,2})$ in terms of $\boldsymbol{Y}_{k1}^{(n)}$ and $\boldsymbol{Y}_{k2}^{(n)}$:

$$\mathrm{E}\big[n\underset{\sim}{H}_{n,2}\big] = \frac{n}{(n-1)^2}\mathrm{E}\big[g_1^{(n)}(\boldsymbol{Y}_{11}^{(n)}, \boldsymbol{Y}_{11}^{(n)})\big]\mathrm{E}\big[g_2^{(n)}(\boldsymbol{Y}_{21}^{(n)}, \boldsymbol{Y}_{21}^{(n)})\big],$$

$$
\begin{aligned}
\mathrm{Var}(n\underset{\sim}{H}_{n,2}) = {}& \frac{4n^2}{(n-1)^3(n-2)^2} \mathrm{Var}\left[g_1^{(n)}(\boldsymbol{Y}_{11}^{(n)}, \boldsymbol{Y}_{11}^{(n)})\right] \mathrm{Var}\left[g_2^{(n)}(\boldsymbol{Y}_{21}^{(n)}, \boldsymbol{Y}_{21}^{(n)})\right] \\
& + \frac{2n}{n-3} \mathrm{Var}\left[g_1^{(n)}(\boldsymbol{Y}_{11}^{(n)}, \boldsymbol{Y}_{12}^{(n)}) + \frac{g_1^{(n)}(\boldsymbol{Y}_{11}^{(n)}, \boldsymbol{Y}_{11}^{(n)})}{n-2} + \frac{g_1^{(n)}(\boldsymbol{Y}_{12}^{(n)}, \boldsymbol{Y}_{12}^{(n)})}{n-2}\right] \\
& \qquad \times \mathrm{Var}\left[g_2^{(n)}(\boldsymbol{Y}_{21}^{(n)}, \boldsymbol{Y}_{22}^{(n)}) + \frac{g_2^{(n)}(\boldsymbol{Y}_{21}^{(n)}, \boldsymbol{Y}_{21}^{(n)})}{n-2} + \frac{g_2^{(n)}(\boldsymbol{Y}_{22}^{(n)}, \boldsymbol{Y}_{22}^{(n)})}{n-2}\right].
\end{aligned}
$$

Using once again Condition (3.5.6), and by a similar argument as in the proof of (B.1.13), we obtain

$$
\mathrm{E}\big[n\underset{\sim}{H}_{n,2}\big] \to \frac{n}{(n-1)^2}\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{11})\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{21})\big] \to 0, \tag{B.1.20}
$$

$$
\begin{aligned}
\mathrm{Var}(n\underset{\sim}{H}_{n,2}) \to {}& \frac{4n^2}{(n-1)^3(n-2)^2} \mathrm{Var}\left[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{11})\right] \mathrm{Var}\left[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{21})\right] \\
& + \frac{2n}{n-3} \mathrm{Var}\left[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12}) + \frac{g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{11})}{n-2} + \frac{g_1(\boldsymbol{Y}_{12}, \boldsymbol{Y}_{12})}{n-2}\right] \\
& \qquad \times \mathrm{Var}\left[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22}) + \frac{g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{21})}{n-2} + \frac{g_2(\boldsymbol{Y}_{22}, \boldsymbol{Y}_{22})}{n-2}\right] \\
\to {}& 2\mathrm{E}\big[g_1(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12})^2\big]\mathrm{E}\big[g_2(\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22})^2\big].
\end{aligned} \tag{B.1.21}
$$

Combining (B.1.20) and (B.1.21), we deduce that (B.1.19) holds.

Finally, Step II is completed by combining (B.1.8), (B.1.9), and (B.1.19) to deduce (B.1.7).

Step III. Notice that $\sup_{i_1,\ldots,i_m \in [\![m]\!]} \mathrm{E}\big[f_k([\boldsymbol{W}_{ki_\ell}]_{\ell=1}^m)^2\big] < \infty$. Proving that $\mathrm{E}\big[(n\underset{\sim}{H}_{n,\ell})^2\big] = o(1)$ for $\ell = 3, 4, \ldots, m$ goes along the same steps as the proof of Theorem 4.2 in the supplement of Shi et al. (2021a); it is omitted here. The fact that $\mathrm{E}\big[(nH_{n,\ell})^2\big] = o(1)$, $\ell = 3, 4, \ldots, m$ follows directly from the theory of degenerate U-statistics (cf. Equation (7) of Section 1.6 in Lee (1990)). The proof is thus complete. $\square$

B.1.3.3 Proof of Theorem 3.5.2

Proof of Theorem 3.5.2. The proof is similar to that of Theorem 3.5.1. The only difference lies in proving (B.1.13)–(B.1.15) and (B.1.20)–(B.1.21). By the continuous mapping theorem (van der Vaart, 1998, Theorem 2.3) and the Skorokhod construction (Shorack, 2017, Chap. 3, Theorem 5.7(viii)), we can assume, without loss of generality, that $\boldsymbol{W}_{ki}^{(n)} \xrightarrow{\text{a.s.}} \boldsymbol{W}_{ki}$. If Condition (3.5.7) holds, then (B.1.13) immediately follows from the dominated convergence theorem and the definitions of $g_k^{(n)}$ and g_k in (3.5.5) and (3.5.2). The proofs for (B.1.14), (B.1.15), (B.1.20), and (B.1.21) are similar. □

B.1.3.4 Proof of Proposition 3.5.2

B.1.3.4.1 Proof of Proposition 3.5.2 ($h = h_{\text{dCov}^2}$)

Proof of Proposition 3.5.2 ($h = h_{\text{dCov}^2}$). Condition (3.5.1) is obvious. Condition (3.5.4) is satisfied in view of Theorem 5 in Székely et al. (2007). We next verify that condition (3.5.6) is satisfied. To do so, let us first show that $g_k^{(n)}(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) \rightrightarrows g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2})$ for $k = 1, 2$. By definitions (3.5.2) and (3.5.5),

$$
\begin{aligned}
g_k^{(n)}(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) &:= \|\boldsymbol{y}_{k1} - \boldsymbol{y}_{k2}\| - \mathrm{E}\|\boldsymbol{y}_{k1} - \boldsymbol{W}_{k3}^{(n)}\| - \mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{y}_{k2}\| + \mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{W}_{k3}^{(n)}\|, \\
\text{and} \quad g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) &:= \|\boldsymbol{y}_{k1} - \boldsymbol{y}_{k2}\| - \mathrm{E}\|\boldsymbol{y}_{k1} - \boldsymbol{W}_{k3}\| - \mathrm{E}\|\boldsymbol{W}_{k4} - \boldsymbol{y}_{k2}\| + \mathrm{E}\|\boldsymbol{W}_{k4} - \boldsymbol{W}_{k3}\|.
\end{aligned}
$$

Noting that J_k, $k = 1, 2$ are continuous, we can assume, as in the proof of Theorem 3.5.2, that $\boldsymbol{W}_{ki}^{(n)} \xrightarrow{\text{a.s.}} \boldsymbol{W}_{ki}$. Since the scores J_k are square-integrable, we obtain that $\mathrm{E}\|\boldsymbol{W}_{ki}^{(n)}\|^2 \to \mathrm{E}\|\boldsymbol{W}_{ki}\|^2$. Using Vitali's theorem (Shorack, 2017, Chap. 3, Theorem 5.5) yields $\boldsymbol{W}_{ki}^{(n)} \xrightarrow{\mathsf{L}^2} \boldsymbol{W}_{ki}$. Therefore, we obtain

$$
\begin{aligned}
\left|\mathrm{E}\|\boldsymbol{y}_{k1} - \boldsymbol{W}_{k3}^{(n)}\| - \mathrm{E}\|\boldsymbol{y}_{k1} - \boldsymbol{W}_{k3}\|\right| &\le \mathrm{E}\|\boldsymbol{W}_{k3}^{(n)} - \boldsymbol{W}_{k3}\|, \\
\left|\mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{y}_{k2}\| - \mathrm{E}\|\boldsymbol{W}_{k4} - \boldsymbol{y}_{k2}\|\right| &\le \mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{W}_{k4}\|,
\end{aligned}
$$

$$\left|\mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{W}_{k3}^{(n)}\| - \mathrm{E}\|\boldsymbol{W}_{k4} - \boldsymbol{W}_{k3}\right| \leq \mathrm{E}\|\boldsymbol{W}_{k3}^{(n)} - \boldsymbol{W}_{k3}\| + \mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{W}_{k4}\|,$$

and, furthermore,

$$\left|g_k^{(n)}(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) - g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2})\right| \leq 2\Big(\mathrm{E}\|\boldsymbol{W}_{k3}^{(n)} - \boldsymbol{W}_{k3}\| + \mathrm{E}\|\boldsymbol{W}_{k4}^{(n)} - \boldsymbol{W}_{k4}\|\Big).$$

The uniform convergence $g_k^{(n)}(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}) \rightrightarrows g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2})$ follows. It is obvious that $g_k(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2})$ is Lipschitz-continuous, and $\mathrm{E}\big[f_k(\boldsymbol{W}_{ki_1}, \ldots, \boldsymbol{W}_{ki_4})^2\big] < \infty$ for all $i_1, \ldots, i_4 \in [\![4]\!]$ as long as J_1, J_2 are weakly regular. □

B.1.3.4.2 Proof of Proposition 3.5.2 ($h = h_M,\ h_D,\ h_R,\ h_{\tau^*}$)

Proof of Proposition 3.5.2 ($h = h_M,\ h_D,\ h_R,\ h_{\tau^*}$). Condition (3.5.1) is obvious. Condition (3.5.4) is satisfied for h_D by Theorem 3(i) in Zhu et al. (2017). For $h = h_M,\ h_R,\ h_{\tau^*}$, we can prove condition (3.5.4) holds as well in a similar way. It is clear that condition (3.5.7) is satisfied for all these four kernel functions. □

B.1.3.5 Proof of Corollary 3.5.1

Proof of Corollary 3.5.1. Combining Proposition 3.5.1 and Theorem 3.5.1, one immediately obtains the limiting null distribution of the rank-based statistic $\underset{\sim}{W}_{\mu}^{(n)}$. □

B.1.3.6 Proof of Proposition 3.5.3

Proof of Proposition 3.5.3. Validity is a direct corollary of Corollary 3.5.1. Uniform validity then follows from validity and exact distribution-freeness. For any fixed alternative in $\mathcal{P}_{d_1,d_2,\infty}^{\mathrm{ac}}$, it holds that $\underset{\sim}{W}_{\mu}^{(n)} \xrightarrow{\text{a.s.}} \mu_{\pm}(\boldsymbol{X}_1, \boldsymbol{X}_2) > 0$ as $n \to \infty$. Thus, $n\,\underset{\sim}{W}_{\mu}^{(n)} \xrightarrow{\text{a.s.}} \infty$ and the result follows. □

B.1.3.7 Proof of Theorem 3.5.3

Proof of Theorem 3.5.3. Let $\boldsymbol{X}_{ni}^*$ and $\boldsymbol{X}_{ni}$, $i \in [\![n]\!]$ be independent copies of $\boldsymbol{X}^*$ and $\boldsymbol{X}$ with $\delta = \delta^{(n)}$, respectively. Let $\mathrm{P}^{(n)} := \otimes_{i=1}^n \mathrm{P}_i^{(n)}$, $\mathrm{Q}^{(n)} := \otimes_{i=1}^n \mathrm{Q}_i^{(n)}$, where $\mathrm{P}_i^{(n)}$ and $\mathrm{Q}_i^{(n)}$ are the distributions of $\boldsymbol{X}_{ni}^*$ and $\boldsymbol{X}_{ni}$, respectively. Define

$$\Lambda^{(n)} := \log \frac{\mathrm{dQ}^{(n)}}{\mathrm{dP}^{(n)}} = \sum_{i=1}^n \log \frac{q_{\boldsymbol{X}}(\boldsymbol{X}_{ni}^*; \delta^{(n)})}{q_{\boldsymbol{X}}(\boldsymbol{X}_{ni}^*; 0)} \quad \text{and} \quad T^{(n)} := \delta^{(n)} \sum_{i=1}^n \dot{\ell}(\boldsymbol{X}_{ni}^*; 0).$$

We proceed in three steps. First, we clarify that $\mathrm{Q}^{(n)}$ is contiguous to $\mathrm{P}^{(n)}$ in order for Le Cam's third lemma (van der Vaart, 1998, Theorem 6.6) to be applicable. Next, we derive the joint limiting null distribution of $(n\underset{\sim}{W}_{\mu}^{(n)}, \Lambda^{(n)})^\top$. Lastly, we employ Le Cam's third lemma to obtain the asymptotic distribution of $(n\underset{\sim}{W}_{\mu}^{(n)}, \Lambda^{(n)})^\top$ under contiguous alternatives.

Step I. In view of Lehmann and Romano (2005, Example 12.3.7), Assumption 3.5.1 entails the contiguity $\mathrm{Q}^{(n)} \triangleleft \mathrm{P}^{(n)}$.

Step II. Next, we derive the limiting joint distribution of $(n\underset{\sim}{W}_{\mu}^{(n)}, \Lambda^{(n)})^\top$ under the null hypothesis. To this end, we first obtain the limiting null distribution of $(nH_{n,2}, T^{(n)})^\top$, where $H_{n,2}$ is defined in (B.1.6). By condition (3.5.1), we write

$$H_{n,2} = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{v=1}^{\infty} \lambda_v \psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i}) \psi_v(\mathbf{Y}_{1j}, \mathbf{Y}_{2j}),$$

where ψ_v is the normalized eigenfunction associated with λ_v and $\mathbf{Y}_{ki} = \mathbf{G}_{k,\pm}^*(\boldsymbol{X}_{ki}^*)$ for $k = 1, 2$. For each positive integer K, consider the "truncated" U-statistic

$$H_{n,2,K} := \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{v=1}^{K} \lambda_v \psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i}) \psi_v(\mathbf{Y}_{1j}, \mathbf{Y}_{2j}).$$

Note that $nH_{n,2}$ and $nH_{n,2,K}$ can be written as

$$nH_{n,2} = \frac{n}{n-1} \Big\{ \sum_{v=1}^{\infty} \lambda_v \Big(\sum_{i=1}^n \frac{\psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})}{\sqrt{n}} \Big)^2 - \sum_{v=1}^{\infty} \lambda_v \Big(\frac{\sum_{i=1}^n \{\psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})\}^2}{n} \Big) \Big\},$$

$$nH_{n,2,K} = \frac{n}{n-1}\Big\{ \sum_{v=1}^{K} \lambda_v \Big(\sum_{i=1}^{n} \frac{\psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})}{\sqrt{n}} \Big)^2 - \sum_{v=1}^{K} \lambda_v \Big(\frac{\sum_{i=1}^{n} \{\psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})\}^2}{n} \Big) \Big\}.$$

To obtain the limiting null distribution of $(nH_{n,2}, T^{(n)})^\top$, first consider the limiting null distribution, for fixed K, of $(nH_{n,2,K}, T^{(n)})^\top$. Let $S_{n,v}$ be a shorthand for $n^{-1/2} \sum_{i=1}^{n} \psi_v(\mathbf{Y}_{1i}, \mathbf{Y}_{2i})$ and observe that

$$\mathrm{E}[S_{n,v}] = \mathrm{E}[T^{(n)}] = 0, \quad \mathrm{Var}[S_{n,v}] = 1, \quad \mathrm{Var}[T^{(n)}] = \mathcal{I}_{\mathbf{X}}(0), \quad \text{and} \quad \mathrm{Cov}[S_{n,v}, T^{(n)}] \to \gamma_v \delta_0.$$

where $\gamma_v := \mathrm{Cov}\left[\psi_v(\mathbf{Y}_1, \mathbf{Y}_2), \dot{\ell}\big((\mathbf{G}_{1,\pm}^{-1}(\mathbf{Y}_1), \mathbf{G}_{2,\pm}^{-1}(\mathbf{Y}_2)); 0\big)\right]$. There exists at least one $v \geq 1$ such that $\gamma_v \neq 0$. Indeed, applying Lemma 4.2 in Nandy et al. (2016) yields

$$\big\{\psi_v(\mathbf{y})\big\}_{v \in \mathbb{Z}_{>0}} = \big\{\psi_{1,v_1}(\mathbf{y}_1)\psi_{2,v_2}(\mathbf{y}_2)\big\}_{v_1, v_2 \in \mathbb{Z}_{>0}},$$

where $\psi_{k,v}(\mathbf{y}_k), v \in \mathbb{Z}_{>0}$ are eigenfunctions associated with the non-zero eigenvalues of the integral equations $\mathrm{E}[g_k(\mathbf{y}_k, \mathbf{W}_{k2})\psi_k(\mathbf{W}_{k2})] = \lambda_k \psi_k(\mathbf{y}_k)$ for $k = 1, 2$. Since $\big\{\psi_{1,v_1}(\mathbf{y}_1)\psi_{2,v_2}(\mathbf{y}_2)\big\}_{v_1, v_2 \in \mathbb{Z}_{\geq 0}}$ (where $\psi_{k,v}(\mathbf{y}_k) := 1$ for $v = 0$, $k = 1, 2$) forms a complete orthogonal basis of the set of square integrable functions, $\gamma_v = 0$ for all $v \geq 1$ thus entails that $\dot{\ell}(\mathbf{x}; 0)$ is additively separable, which contradicts Assumption 3.5.1(iv). Therefore, $\gamma_{v^*} \neq 0$ for some v^*. Applying the multivariate central limit theorem (Bhattacharya and Ranga Rao, 1986, Equation (18.24)), we deduce

$$(S_{n,1}, \ldots, S_{n,K}, T^{(n)})^\top \overset{\mathrm{P}^{(n)}}{\rightsquigarrow} (\xi_1, \ldots, \xi_K, V_K)^\top \sim N_{K+1}\Bigg(\begin{pmatrix} \mathbf{0}_K \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{I}_p & \delta_0 \mathbf{v} \\ \delta_0 \mathbf{v}^\top & \delta_0^2 \mathcal{I} \end{pmatrix} \Bigg),$$

where $\mathcal{I} := \mathcal{I}_{\mathbf{X}}(0)$ and $\mathbf{v} = (\gamma_1, \ldots, \gamma_K)^\top$. Thus, V_K can be expressed as

$$\Big(\delta_0^2 \mathcal{I}\Big)^{1/2} \Big\{ \sum_{v=1}^{K} c_v \xi_v + \Big(1 - \sum_{v=1}^{K} c_v^2\Big)^{1/2} \xi_0 \Big\}$$

where $c_v := \mathcal{I}^{-1/2}\gamma_v$, and ξ_0 is standard Gaussian, independent of $\xi_1, \ldots, \xi_K$. Then, by the continuous mapping theorem (van der Vaart, 1998, Theorem 2.3) and Slutsky's theorem (van der Vaart, 1998, Theorem 2.8),

$$(nH_{n,2,K}, T^{(n)})^\top \overset{\mathrm{P}^{(n)}}{\rightsquigarrow} \left(\sum_{v=1}^{K} \lambda_v(\xi_v^2 - 1), \left(\delta_0^2 \mathcal{I}\right)^{1/2} \left\{ \sum_{v=1}^{K} c_v \xi_v + \left(1 - \sum_{v=1}^{K} c_v^2\right)^{1/2} \xi_0 \right\} \right)^\top \tag{B.1.22}$$

for any K. This entails

$$(nH_{n,2}, T^{(n)})^\top \overset{\mathrm{P}^{(n)}}{\rightsquigarrow} \left(\sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1), \left(\delta_0^2 \mathcal{I}\right)^{1/2} \left\{ \sum_{v=1}^{\infty} c_v \xi_v + \left(1 - \sum_{v=1}^{\infty} c_v^2\right)^{1/2} \xi_0 \right\} \right)^\top. \tag{B.1.23}$$

Indeed, putting

$$M_K := \sum_{v=1}^{K} \lambda_v(\xi_v^2 - 1), \qquad V_K := \left(\delta_0^2 \mathcal{I}\right)^{1/2} \left\{ \sum_{v=1}^{K} c_v \xi_v + \left(1 - \sum_{v=1}^{K} c_v^2\right)^{1/2} \xi_0 \right\},$$
$$M := \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1), \quad \text{and} \quad V := \left(\delta_0^2 \mathcal{I}\right)^{1/2} \left\{ \sum_{v=1}^{\infty} c_v \xi_v + \left(1 - \sum_{v=1}^{\infty} c_v^2\right)^{1/2} \xi_0 \right\},$$

it suffices, in order to to prove (B.1.23), to show that, for any $a, b \in \mathbb{R}$,

$$\left| \mathrm{E}\left[\exp\left\{ \mathrm{i}anH_{n,2} + \mathrm{i}bT^{(n)} \right\} \right] - \mathrm{E}\left[\exp\left\{ \mathrm{i}aM + \mathrm{i}bV \right\} \right] \right| \to 0 \quad \text{as } n \to \infty. \tag{B.1.24}$$

We have

$$\begin{aligned}
&\left| \mathrm{E}\left[\exp\left\{ \mathrm{i}anH_{n,2} + \mathrm{i}bT^{(n)} \right\} \right] - \mathrm{E}\left[\exp\left\{ \mathrm{i}aM + \mathrm{i}bV \right\} \right] \right| \\
\leq &\left| \mathrm{E}\left[\exp\left\{ \mathrm{i}anH_{n,2} + \mathrm{i}bT^{(n)} \right\} \right] - \mathrm{E}\left[\exp\left\{ \mathrm{i}anH_{n,2,K} + \mathrm{i}bT^{(n)} \right\} \right] \right| \\
&+ \left| \mathrm{E}\left[\exp\left\{ \mathrm{i}anH_{n,2,K} + \mathrm{i}bT^{(n)} \right\} \right] - \mathrm{E}\left[\exp\left\{ \mathrm{i}aM_K + \mathrm{i}bV_K \right\} \right] \right| \\
&+ \left| \mathrm{E}\left[\exp\left\{ \mathrm{i}aM_K + \mathrm{i}bV_K \right\} \right] - \mathrm{E}\left[\exp\left\{ \mathrm{i}aM + \mathrm{i}bV \right\} \right] \right| =: I + II + III, \quad \text{say,}
\end{aligned}$$

where it follows from page 82 of Lee (1990) and Equation (4.3.10) in Koroljuk and Borovskich

(1994) that

$$I \leq \mathrm{E}\Big|\exp\Big\{\mathrm{i}an(H_{n,2} - H_{n,2,K})\Big\} - 1\Big| \leq \Big\{\mathrm{E}\Big|an(H_{n,2} - H_{n,2,K})\Big|^2\Big\}^{1/2} = \Big\{\frac{2na^2}{n-1}\sum_{v=K+1}^{\infty}\lambda_v^2\Big\}^{1/2}$$

and

$$\begin{aligned} III &\leq \mathrm{E}\Big|\exp\Big\{\mathrm{i}a(M_K - M) + \mathrm{i}b(V_K - V)\Big\} - 1\Big| \leq \Big\{\mathrm{E}\Big|a(M_K - M) + b(V_K - V)\Big|^2\Big\}^{1/2} \\ &\leq \Big\{2\Big(2a^2\sum_{v=K+1}^{\infty}\lambda_v^2 + 2b^2\delta_0^2\mathcal{I}\sum_{v=K+1}^{\infty}c_v^2\Big)\Big\}^{1/2}. \end{aligned}$$

Since by condition (3.5.1)

$$\sum_{v=1}^{\infty}\lambda_v^2 = \mathrm{Var}(g_1(\boldsymbol{W}_{11}, \boldsymbol{W}_{12})) \cdot \mathrm{Var}(g_2(\boldsymbol{W}_{21}, \boldsymbol{W}_{22})) \in (0, \infty) \quad \text{and} \quad \sum_{v=1}^{\infty}c_v^2 = \mathcal{I}^{-1}\sum_{v=1}^{\infty}\gamma_v^2 \leq 1,$$

we conclude that, for any $\epsilon > 0$, there exists K_0 such that $I < \epsilon/3$ and $III < \epsilon/3$ for all n and all $K \geq K_0$. For this K_0, we also have, by (B.1.22), that $II < \epsilon/3$ for all n sufficiently large; (B.1.24), hence (B.1.23), follow.

Now, as in van der Vaart (1998, Theorem 7.2),

$$\Lambda^{(n)} - T^{(n)} + \delta_0^2\mathcal{I}/2 \xrightarrow{\mathrm{P}^{(n)}} 0. \tag{B.1.25}$$

Combining (B.1.23) and (B.1.25) yields

$$(nH_{n,2}, \Lambda^{(n)})^\top \overset{\mathrm{P}^{(n)}}{\rightsquigarrow} \left(\sum_{v=1}^{\infty}\lambda_v(\xi_v^2 - 1), \Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big\{\sum_{v=1}^{\infty}c_v\xi_v + \Big(1 - \sum_{v=1}^{\infty}c_v^2\Big)^{1/2}\xi_0\Big\} - \frac{\delta_0^2\mathcal{I}}{2}\right)^\top. \tag{B.1.26}$$

Equation (1.6.7) in Lee (1990, p. 30), along with the fact that $H_{n,1} = 0$, implies that $(n\underset{\sim}{W}_{\mu}^{(n)}, \Lambda^{(n)})^\top$ has the same limiting distribution as (B.1.26) under $\mathrm{P}^{(n)}$.

Step III. Finally we employ the general form (van der Vaart, 1998, Theorem 6.6) of Le

Cam's third lemma, which by condition (3.5.4) entails

$$
\begin{aligned}
&\mathrm{Q}^{(n)}(n\underset{\sim}{W}_{\mu}^{(n)} \leq q_{1-\alpha}) \\
&\to \mathrm{E}\Big[\mathbb{1}\Big(\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)\leq q_{1-\alpha}\Big)\cdot\exp\Big\{\Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big(\sum_{v=1}^{\infty}c_v\xi_v+\Big(1-\sum_{v=1}^{\infty}c_v^2\Big)^{1/2}\xi_0\Big)-\frac{\delta_0^2\mathcal{I}}{2}\Big\}\Big] \\
&\leq \mathrm{E}\Big[\mathbb{1}\Big\{\Big|\xi_{v^*}\Big|\leq\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\Big\}\cdot\exp\Big\{\Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big(\sum_{v=1}^{\infty}c_v\xi_v+\Big(1-\sum_{v=1}^{\infty}c_v^2\Big)^{1/2}\xi_0\Big)-\frac{\delta_0^2\mathcal{I}}{2}\Big\}\Big] \\
&= \mathrm{E}\Big[\mathbb{1}\Big\{\Big|\xi_{v^*}\Big|\leq\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\Big\}\cdot\exp\Big\{\Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big(c_{v^*}\xi_{v^*}+\Big(1-c_{v^*}^2\Big)^{1/2}\xi_0\Big)-\frac{\delta_0^2\mathcal{I}}{2}\Big\}\Big] \\
&= \Phi\Big(\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}-c_{v^*}\Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big)-\Phi\Big(-\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}-c_{v^*}\Big(\delta_0^2\mathcal{I}\Big)^{1/2}\Big) \\
&\leq 2\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\varphi\Big(\Big\{|c_{v^*}|\cdot\Big(\delta_0^2\mathcal{I}\Big)^{1/2}-\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\Big\}_+\Big),
\end{aligned}
$$

a quantity which is arbitrarily small for large enough δ_0, irrespective of the sign of c_{v^*}. □

B.1.3.8 Proof of Theorem 3.5.4

Proof of Theorem 3.5.4. This result is a standard result connecting the Fisher information to the usual lower bound of rate $n^{-1/2}$ (Groeneboom and Jongbloed, 2014, Chap. 6). Recall that $\boldsymbol{X}_{ni}^*$ and $\boldsymbol{X}_{ni}$, $i \in [\![n]\!]$ are independent copies of $\boldsymbol{X}^*$ and $\boldsymbol{X}$, respectively, with $\delta = \delta^{(n)} = n^{-1/2}\delta_0$. Recall $\mathrm{P}^{(n)} := \otimes_{i=1}^n \mathrm{P}_i^{(n)}$, $\mathrm{Q}^{(n)} := \otimes_{i=1}^n \mathrm{Q}_i^{(n)}$, where $\mathrm{P}_i^{(n)}$ and $\mathrm{Q}_i^{(n)}$ are the distributions of $\boldsymbol{X}_{ni}^*$ and $\boldsymbol{X}_{ni}$, respectively. It suffices to prove that for any small $0 < \beta < 1-\alpha$, there exists $|\delta_0| = c_\beta$ such that, for all sufficiently large n, $\mathrm{TV}(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)}) < \beta$, which is implied by $\mathrm{HL}(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)}) < \beta$ using the fact that total variation and Hellinger distances satisfy

$$\mathrm{TV}(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)}) \leq \mathrm{HL}(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)})$$

(Tsybakov, 2009, Equation (2.20)). It is also known (Tsybakov, 2009, p. 83) that

$$1 - \frac{\mathrm{HL}^2(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)})}{2} = \prod_{i=1}^n \Big(1 - \frac{\mathrm{HL}^2(\mathrm{Q}_i^{(n)}, \mathrm{P}_i^{(n)})}{2}\Big).$$

Lehmann and Romano (2005, Example 13.1.1) shows that, under Assumption 3.5.1,

$$n \times \mathrm{HL}^2(\mathrm{Q}_i^{(n)}, \mathrm{P}_i^{(n)}) \to \frac{\delta_0^2 \mathcal{I}_{\boldsymbol{X}}(0)}{4};$$

notice that here the definition of $\mathrm{HL}^2(\mathrm{Q}, \mathrm{P})$ differs with that in Lehmann and Romano (2005, Definition 13.1.3) by a factor of 2. Therefore,

$$1 - \frac{\mathrm{HL}^2(\mathrm{Q}^{(n)}, \mathrm{P}^{(n)})}{2} \longrightarrow \exp\Big\{ - \frac{\delta_0^2 \mathcal{I}_{\boldsymbol{X}}(0)}{8} \Big\}.$$

The desired result follows by taking $c_\beta > 0$ such that

$$\exp\Big\{ - \frac{c_\beta^2 \mathcal{I}_{\boldsymbol{X}}(0)}{8} \Big\} = 1 - \frac{\beta^2}{8}.$$

This completes the proof. □

B.1.3.9 Proof of Example 3.5.1

B.1.3.9.1 Proof of Example 3.5.1(i)

Proof of Example 3.5.1(i). We need to verify Assumption 3.5.2. Items (i) and (ii) are obvious. For (iii), following the proof of Lemma 3.2.1 in Gieser (1993), when $\boldsymbol{X}_1^*$ and $\boldsymbol{X}_2^*$ are elliptically symmetric with parameters $\mathbf{0}_{d_1}$, $\mathbf{\Sigma}_1$ and $\mathbf{0}_{d_2}$, $\mathbf{\Sigma}_2$, respectively, we obtain

$$\dot{\ell}(\boldsymbol{x}; 0) = -2(\mathbf{M}_1 \boldsymbol{x}_2)^\top \mathbf{\Sigma}_1^{-1} \boldsymbol{x}_1 \cdot \rho_1\Big(\boldsymbol{x}_1^\top \mathbf{\Sigma}_1^{-1} \boldsymbol{x}_1\Big) - 2(\mathbf{M}_2 \boldsymbol{x}_1)^\top \mathbf{\Sigma}_2^{-1} \boldsymbol{x}_2 \cdot \rho_2\Big(\boldsymbol{x}_2^\top \mathbf{\Sigma}_2^{-1} \boldsymbol{x}_2\Big).$$

Consequently, the condition that $\mathrm{E}\,[\|\boldsymbol{Z}_k^*\|^2 \rho_k(\|\boldsymbol{Z}_k^*\|^2)^2] < \infty$ for $k = 1, 2$ is sufficient for $\mathcal{I}_{\boldsymbol{X}}(0) = \mathrm{E}\big[\dot{\ell}(\boldsymbol{X}; 0)^2\big] < \infty$. If $\mathcal{I}_{\boldsymbol{X}}(0) = 0$, then we must have

$$\rho_1\Big(\boldsymbol{x}_1^\top \mathbf{\Sigma}_1^{-1} \boldsymbol{x}_1\Big) = \rho_2\Big(\boldsymbol{x}_2^\top \mathbf{\Sigma}_2^{-1} \boldsymbol{x}_2\Big) = C_\rho$$

for some constant $C_\rho \neq 0$ and

$$(\mathbf{M}_1 \boldsymbol{x}_2)^\top \mathbf{\Sigma}_1^{-1} \boldsymbol{x}_1 + (\mathbf{M}_2 \boldsymbol{x}_1)^\top \mathbf{\Sigma}_2^{-1} \boldsymbol{x}_2 = \boldsymbol{x}_1^\top \mathbf{\Sigma}_1^{-1} (\mathbf{M}_1 \mathbf{\Sigma}_2 + \mathbf{\Sigma}_1 \mathbf{M}_2^\top) \mathbf{\Sigma}_2^{-1} \boldsymbol{x}_2 = 0$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2$. This contradicts the assumption that $\boldsymbol{\Sigma}_1 \mathbf{M}_2^\top + \mathbf{M}_1 \boldsymbol{\Sigma}_2 \neq \mathbf{0}$ and completes the proof. □

B.1.3.9.2 Proof of Example 3.5.1(ii)

Proof of Example 3.5.1(ii). For the multivariate normal, $\phi_k(t) = \exp(-t/2)$ and $\rho_k(t) = -1/2$, so that all conditions in Example 3.5.1(i) are satisfied. For a multivariate t-distribution with ν_k degrees of freedom,

$$\phi_k(t) = (1 + t/\nu_k)^{-(\nu_k + d_k)/2} \quad \text{and} \quad \rho_k(t) = -2^{-1}(1 + d_k/\nu_k)(1 + t/\nu_k)^{-1}.$$

It is easily checked that all conditions in Example 3.5.1(i) are satisfied when $\nu_k > 2$; see Gieser (1993, p. 44–46). □

B.1.3.10 Proof of Example 3.5.2

Proof of Example 3.5.2. Since q^* is continuous and has compact support, it is upper bounded by some constant, say $C_q > 1$, and then Assumption 3.5.3(i) holds with $\delta^* = C_q^{-1}$. The rest of Assumption 3.5.3 can be easily verified. □

B.1.3.11 Proof of Proposition 3.5.4

Proof of Propositiion 3.5.4. (1) Konijn family. It is clear that Assumption 3.5.1(i),(iii) is satisfied. Gieser (1993, Appendix B, p. 105–107) shows that Assumption 3.5.2 implies Assumption 3.5.1(ii). To verify Assumption 3.5.1(iv), notice that

$$\dot{\ell}(\boldsymbol{x}; 0) = -2(\mathbf{M}_1 \boldsymbol{x}_2)^\top \Big(\nabla q_1(\boldsymbol{x}_1)/q_1(\boldsymbol{x}_1) \Big) - 2(\mathbf{M}_2 \boldsymbol{x}_1)^\top \Big(\nabla q_2(\boldsymbol{x}_2)/q_2(\boldsymbol{x}_2) \Big)$$

following the proof of Lemma 3.2.1 in Gieser (1993).

(2) Mixture family. Direct computation yields that

$$\dot{\ell}(\boldsymbol{x};\delta) = \frac{q^*(\boldsymbol{x}) - q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)}{(1-\delta)\{q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)\} + \delta q^*(\boldsymbol{x})}.$$

The rest directly follows from Theorem 12.2.1 in Lehmann and Romano (2005). □

B.2 Auxiliary results

B.2.1 Auxiliary results for Section 3.2

The concept of GSC unifies a surprisingly large number of well-known dependence measures. Moreover, only two types of subgroups are needed, namely, $H_\tau^m := \langle (1\ 2)\rangle = \{(1), (1\ 2)\} \subseteq \mathfrak{S}_m$ for $m = 2$ and $H_*^m := \langle (1\ 4), (2\ 3)\rangle = \{(1), (1\ 4), (2\ 3), (1\ 4)(2\ 3)\} \subseteq \mathfrak{S}_m$ for $m \geq 4$. The following result illustrates this fact with four classical examples of univariate dependence measures, namely, the tau of Kendall (1938), the D of Hoeffding (1948), the R of Blum et al. (1961), and the τ^* of Bergsma and Dassios (2014) which, as shown by Drton et al. (2020), is connected to the work of Yanagimoto (1970). Below, we write $\boldsymbol{w} = (w_1, \ldots, w_m) \mapsto f_k(\boldsymbol{w})$, $k = 1, 2$ for the kernel functions of an mth order univariate GSC; note that not all components of $\boldsymbol{w}$ need to have an impact on $f_k(\boldsymbol{w})$: see, for instance the kernel f_1 of the 6th order Blum–Kiefer–Rosenblatt GSC, which is mapping $\boldsymbol{w} = (w_1, \ldots, w_6)$ to $\mathbb{R}_{\geq 0}$ but does not depend on w_6 (f_2 does).

Example B.2.1 (Examples of univariate GSCs)**.**

(a) Kendall's tau is a 2nd order GSC with $H = H_\tau^2$ and

$$f_1(\boldsymbol{w}) = f_2(\boldsymbol{w}) = \mathbb{1}(w_1 < w_2) \text{ on } \mathbb{R}^2,$$

which can be proved as follows:

$$\mu_{f_1,f_2,H_\tau^2}(X_1, X_2) := \mathrm{E}[k_{f_1,f_2,H_\tau^2}((X_{11}, X_{21}), (X_{12}, X_{22}))]$$

$$= \mathrm{E}[\{\mathbb{1}(X_{11} < X_{12}) - \mathbb{1}(X_{12} < X_{11})\}\{\mathbb{1}(X_{21} < X_{22}) - \mathbb{1}(X_{22} < X_{21})\}]$$
$$= \mathrm{E}[\mathrm{sign}(X_{11} - X_{12})\,\mathrm{sign}(X_{21} - X_{22})] =: \tau;$$

see also Example 5 in Lee (1990, Chapter 1.2) for the last expression of Kendall's τ;

(b) Hoeffding's D is a 5th order GSC with $H = H_*^5$ and

$$f_1(\boldsymbol{w}) = f_2(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_5) \quad \text{on} \quad \mathbb{R}^5;$$

(c) Blum–Kiefer–Rosenblatt's R is a 6th order GSC with $H = H_*^6$ and

$$f_1(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_5), \qquad f_2(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_6) \quad \text{on} \quad \mathbb{R}^6;$$

(d) Bergsma–Dassios–Yanagimoto's τ^* is a 4th order GSC with $H = H_*^4$ and

$$f_1(\boldsymbol{w}) = f_2(\boldsymbol{w}) = \mathbb{1}(\max\{w_1, w_2\} < \min\{w_3, w_4\}) \text{ on } \mathbb{R}^4.$$

Remark B.2.1. Distinct choices of the kernels f_1 and f_2 do not necessarily imply distinct GSCs. For example, Weihs et al. (2018, Proposition 1(ii)) showed that Hoeffding's D in Example B.2.1(b) is a 5th order GSC with $H = H_*^5$ also for

$$f_1(\boldsymbol{w}) = f_2(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_5 < \max\{w_3, w_4\}) \text{ on } \mathbb{R}^5;$$

similarly, for Blum–Kiefer–Rosenblatt's R, the kernels in Example B.2.1(c) can be replaced with

$$f_1(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_5 < \max\{w_3, w_4\}) \text{ on } \mathbb{R}^6,$$
$$f_2(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\max\{w_1, w_2\} \le w_6 < \max\{w_3, w_4\}) \text{ on } \mathbb{R}^6.$$

B.2.2 Auxiliary results for Section 3.3

The next proposition collects several properties of center-outward distribution functions.

Proposition B.2.1. *Let* $\mathbf{F}_\pm$ *be the center-outward distribution function of* $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$*. Then,*

(i) (Hallin, 2017, Proposition 4.2(i), Hallin et al., 2021a, Proposition 2.1(i),(iii)) $\mathbf{F}_\pm$ *is a probability integral transformation of* $\mathbb{R}^d$*, namely,* $\boldsymbol{Z} \sim \mathrm{P}$ *if and only if* $\mathbf{F}_\pm(\boldsymbol{Z}) \sim \mathrm{U}_d$*;*

(ii) (Hallin et al., 2021a, Proposition 2.1(ii)) if $\boldsymbol{Z} \sim \mathrm{P}$, $\|\mathbf{F}_\pm(\boldsymbol{Z})\|$ *is uniform over* $[0, 1)$*,* $\mathbf{F}_\pm(\boldsymbol{Z})/\|\mathbf{F}_\pm(\boldsymbol{Z})\|$ *is uniform over the sphere* $\mathcal{S}_{d-1}$*, and they are mutually independent.*

Writing $\mathbf{F}_\pm^{\boldsymbol{Z}}$ *for the center-outward distribution function of* $\boldsymbol{Z} \sim \mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$*,*

(iii) (Hallin et al., 2020, Proposition 2.2) for any $\boldsymbol{v} \in \mathbb{R}^d, a \in \mathbb{R}_{>0}$*, and orthogonal* $d \times d$ *matrix* $\mathbf{O}$*,*

$$\mathbf{F}_\pm^{\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}}(\boldsymbol{v} + a\mathbf{O}\boldsymbol{z}) = \mathbf{O}\mathbf{F}_\pm^{\boldsymbol{Z}}(\boldsymbol{z}) \text{ for all } \boldsymbol{z} \in \mathbb{R}^d.$$

Letting $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ *be independent copies of* $\boldsymbol{Z} \sim \mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ *with center-outward distribution function* $\mathbf{F}_\pm$*,*

(iv) (Hallin, 2017, Proposition 6.1(ii), Hallin et al., 2021a, Proposition 2.5(ii)) for any decomposition n_0, n_R, n_S *of* n*, the random vector* $[\mathbf{F}_\pm^{(n)}(\boldsymbol{Z}_1), \ldots, \mathbf{F}_\pm^{(n)}(\boldsymbol{Z}_n)]$ *is uniformly distributed over all distinct arrangements of the grid* $\mathfrak{G}_{\boldsymbol{n}}^d$*;*

(v) (del Barrio et al., 2018, Proof of Theorem 3.1, Hallin et al., 2021a, Proof of Proposition 3.3) as n_R *and* $n_S \to \infty$*, for every* $i \in [\![n]\!]$*,*

$$\left\|\mathbf{F}_\pm^{(n)}(\boldsymbol{Z}_i) - \mathbf{F}_\pm(\boldsymbol{Z}_i)\right\| \xrightarrow{\text{a.s.}} 0.$$

Proof of Proposition B.2.1. We give an independent proof of part (iii). In view of Definition 3.3.1, there exists a convex function Ψ such that $\mathbf{F}_\pm^{\boldsymbol{Z}} = \nabla\Psi$. It is obvious that $\mathbf{F}_\pm^{\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}}$

defined implicitly by

$$\mathbf{F}_{\pm}^{\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}}(\boldsymbol{v}+a\mathbf{O}\boldsymbol{z}) = \mathbf{O}\mathbf{F}_{\pm}^{\boldsymbol{Z}}(\boldsymbol{z}),$$

satisfies (ii) and (iii) in Definition 3.3.1. It only remains, thus, to construct a convex function Ψ^* such that $\mathbf{F}_{\pm}^{\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}} = \nabla\Psi^*$. Noting that $\mathbf{F}_{\pm}^{\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}}(\boldsymbol{z}) = \mathbf{O}\mathbf{F}_{\pm}^{\boldsymbol{Z}}(a^{-1}\mathbf{O}^{-1}(\boldsymbol{z}-\boldsymbol{v}))$, it is easy to check that $\boldsymbol{z} \mapsto \Psi^*(\boldsymbol{z}) := a\Psi\left(a^{-1}\mathbf{O}^{-1}(\boldsymbol{z}-\boldsymbol{v})\right)$ is convex, and thus continuous and almost everywhere differentiable, with $\nabla\Psi^*(\boldsymbol{v}+a\mathbf{O}\boldsymbol{Z}) = \mathbf{O}\nabla\Psi(\boldsymbol{z})$. □

Proposition B.2.2. *(Hallin, 2017, Proposition 5.1, del Barrio et al., 2018, Theorem 3.1, del Barrio et al., 2020, Theorem 2.5, and Hallin et al., 2021a, Proposition 2.3) Consider the following classes of distributions:*

- *the class $\mathcal{P}_d^+$ of distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ with nonvanishing probability density, namely, with Lebesgue density f such that, for all $D > 0$ there exist constants $\lambda_{D;f} < \Lambda_{D;f} \in (0,\infty)$ such that $\lambda_{D;f} \leq f(\boldsymbol{z}) \leq \Lambda_{D;f}$ for all $\|\boldsymbol{z}\| \leq D$;*
- *the class $\mathcal{P}_d^{\mathrm{conv}}$ of distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ with convex support $\overline{\mathrm{supp}}(\mathrm{P})$ and a density that is nonvanishing over this support, namely, with density f such that, for all $D > 0$ there exist constants $\lambda_{D;f} < \Lambda_{D;f} \in (0,\infty)$ such that $\lambda_{D;f} \leq f(\boldsymbol{z}) \leq \Lambda_{D;f}$ for all $\boldsymbol{z} \in \mathrm{supp}(\mathrm{P})$ with $\|\boldsymbol{z}\| \leq D$;*
- *the class $\mathcal{P}_d^{\pm}$ of distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ that are push-forwards of U_d of the form $\mathrm{P} = \nabla\Upsilon\sharp\mathrm{U}_d$ ($\nabla\Upsilon$ the gradient of a convex function) and a homeomorphism from the punctured ball $\mathbb{S}_d\backslash\{\mathbf{0}_d\}$ to $\nabla\Upsilon(\mathbb{S}_d\backslash\{\mathbf{0}_d\})$ such that $\nabla\Upsilon(\{\mathbf{0}_d\})$ is compact, convex, and has Lebesgue measure zero;*
- *the class $\mathcal{P}_d^{\#}$ of all distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ such that, denoting by $\mathbf{F}_{\pm}^{(n)}$ the sample distribution function computed from an n-tuple $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ of independent copies of $\boldsymbol{Z} \sim \mathrm{P}$,*

  $$\max_{1\leq i\leq n}\left\|\mathbf{F}_{\pm}^{(n)}(\boldsymbol{Z}_i) - \mathbf{F}_{\pm}(\boldsymbol{Z}_i)\right\| \xrightarrow{\text{a.s.}} 0$$

  as n_R and $n_S \to \infty$ (a Glivenko-Cantelli property).

It holds that $\mathcal{P}_d^+ \subsetneq \mathcal{P}_d^{\text{conv}} \subsetneq \mathcal{P}_d^{\pm} \subseteq \mathcal{P}_d^{\#} \subsetneq \mathcal{P}_d^{\text{ac}}$.

B.2.3 Auxiliary results for Section 3.4

The time complexity of computing the optimal matching and nearly optimal matchings is summarized in the following proposition.

Proposition B.2.3. *The optimal matching problem* (3.3.1) *yielding* $[\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})]_{i=1}^n$ *and* $[\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})]_{i=1}^n$ *can be solved in* $O(n^3)$ *time via the refined Hungarian algorithm (Dinic and Kronrod, 1969; Tomizawa, 1971; Edmonds and Karp, 1970, 1972). Moreover,*

(i) if we assume that $c_{ij},\ i,j \in [\![n]\!]$ *all are integers and bounded by some (positive) integer* N*, which can be achieved by scaling and rounding, then there exists an optimal matching algorithm solving the problem in* $O(n^{5/2}\log(nN))$ *time (Gabow and Tarjan, 1989);*

(ii) if $d = 2$ *and* $c_{ij},\ i,j \in [\![n]\!]$ *all are integers and bounded by some (positive) integer* N*, there exists an exact an optimal matching algorithm solving the problem in* $O(n^{3/2+\delta}\log(N))$ *time for any arbitrarily small constant* $\delta > 0$ *(Sharathkumar and Agarwal, 2012);*

(iii) if $d \geq 3$*, there is an algorithm computing a* $(1+\epsilon)$*-approximate perfect matching in*

$$O\left(n^{3/2}\epsilon^{-1}\tau(n,\epsilon)\log^4(n/\epsilon)\log\left(\max c_{ij}/\min c_{ij}\right)\right) \qquad \textit{time},$$

where a $(1+\epsilon)$*-approximate perfect matching for* $\epsilon > 0$ *is a bijection* σ *from* $[\![n]\!]$ *to itself such that* $\sum_{i=1}^n c_{i\sigma(i)}$ *is no larger than* $(1+\epsilon)$ *times the cost of the optimal matching and* $\tau(n,\epsilon)$ *is the query and update time of an* ϵ/c*-approximate nearest neighbor data structure for some constant* $c > 1$ *(Agarwal and Sharathkumar, 2014).*

Once $[\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})]_{i=1}^n$ and $[\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})]_{i=1}^n$ are obtained, a naive approach to the computation of $\underset{\sim}{W}^{(n)}$, on the other hand, requires at most a $O(n^m)$ time complexity. Great speedups are possible, however, in particular cases and the next proposition summarizes the results for the various center-outward rank-based statistics listed in Example 3.4.1.

Proposition B.2.4. *Assuming that* $[\mathbf{G}^{(n)}_{1,\pm}(\boldsymbol{X}_{1i})]_{i=1}^n$ *and* $[\mathbf{G}^{(n)}_{2,\pm}(\boldsymbol{X}_{2i})]_{i=1}^n$ *have been previously obtained, one can compute*

(i) $\underset{\sim}{W}^{(n)}_{\text{dCov}}$ *in* $O(n^2)$ *time (Székely and Rizzo, 2013,* Definition 1, *Székely and Rizzo, 2014,* *Definition 2, Proposition 1, Huo and Székely, 2016,* Lemma 3.1*)*

(ii) $\underset{\sim}{W}^{(n)}_{M}$ *in* $O(n(\log n)^{d_1+d_2-1})$ *time (Weihs et al., 2018, p. 557, end of Sec. 5.2),*

(iii) $\underset{\sim}{W}^{(n)}_{D}$ *in* $O(n^3)$ *time (Zhu et al., 2017, Theorem 1),*

(iv) $\underset{\sim}{W}^{(n)}_{R}$ *in* $O(n^4)$ *time as proved in Section A.3.4 of the supplement,*

(v) $\underset{\sim}{W}^{(n)}_{\tau^*}$ *in* $O(n^4)$ *time by definition.*

If, moreover, approximate values are allowed, one can compute

(i) approximate $\underset{\sim}{W}^{(n)}_{\text{dCov}}$ *in* $O(nK\log n)$ *time (Huo and Székely, 2016, Theorem 4.1, Chaudhuri and Hu, 2019, Theorem 3.1),*

(ii) approximate $\underset{\sim}{W}^{(n)}_{D}$ *in* $O(nK\log n)$ *time (Weihs et al., 2018, p. 557),*

(iii) approximate $\underset{\sim}{W}^{(n)}_{R}$ *in* $O(nK\log n)$ *time (Drton et al., 2020,* Equation (6.1)*, Weihs et al., 2018, p. 557, Even-Zohar and Leng, 2021,* Corollary 4*),*

(iv) approximate $\underset{\sim}{W}^{(n)}_{\tau^*}$ *in* $O(nK\log n)$ *time (Even-Zohar and Leng, 2021, Corollary 4).*

These approximations consider random projections to speed up computation; K stands for the number of random projections. See also Huang and Huo (2017, Sec. 3.1).

Proof of Proposition B.2.4. We only illustrate how to efficiently compute U-statistic estimates of Hoeffding's multivariate projection-averaging D and Blum–Kiefer–Rosenblatt's multivariate projection-averaging R. The other claims straightforwardly follow from the sources provided in the proposition.

Zhu et al. (2017) showed how to efficiently compute a V-statistic estimate of Hoeffding's multivariate projection-averaging D. Let us show how to efficiently compute the corresponding U-statistic. We define arrays $(a_{k\ell rs})_{\ell,r,s\in[\![n]\!]}$ for $k=1,2$ as

$$\begin{cases} a_{k\ell rs} := \mathsf{Arc}(\boldsymbol{y}_{k\ell} - \boldsymbol{y}_{ks}, \boldsymbol{y}_{kr} - \boldsymbol{y}_{ks}) & \text{if } [\ell, r, s] \in I_3^n, \\ a_{k\ell rs} := 0 & \text{otherwise.} \end{cases}$$

Their U-centered versions $(A_{k\ell rs})_{\ell,r,s\in[\![n]\!]}$ for $k=1,2$ are

$$A_{k\ell rs} := \begin{cases} a_{k\ell rs} - \dfrac{1}{n-3}\displaystyle\sum_{i=1}^{n} a_{kirs} - \dfrac{1}{n-3}\sum_{j=1}^{n} a_{k\ell js} + \dfrac{1}{(n-2)(n-3)}\sum_{i,j=1}^{n} a_{kijs} & \text{if } [\ell, r, s] \in I_3^n, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\binom{n}{5}^{-1} \sum_{i_1<\cdots<i_5} h_D\Big((\boldsymbol{y}_{1i_1}, \boldsymbol{y}_{2i_1}), \ldots, (\boldsymbol{y}_{1i_5}, \boldsymbol{y}_{2i_5})\Big) = \frac{1}{n(n-1)(n-4)} \sum_{[\ell,r,s]\in I_3^n} A_{1\ell rs} A_{2\ell rs},$$

which clearly has $O(n^3)$ complexity.

Turning to Blum–Kiefer–Rosenblatt's multivariate projection-averaging R, define, for $k=1,2$, the arrays $(b_{k\ell rst})_{\ell,r,s,t\in[\![n]\!]}$ as

$$\begin{cases} b_{1\ell rst} := \mathsf{Arc}(\boldsymbol{y}_{1\ell} - \boldsymbol{y}_{1s}, \boldsymbol{y}_{1r} - \boldsymbol{y}_{1s}) \text{ and } b_{2\ell rst} := \mathsf{Arc}(\boldsymbol{y}_{2\ell} - \boldsymbol{y}_{2t}, \boldsymbol{y}_{2r} - \boldsymbol{y}_{2t}) & \text{if } [\ell, r, s, t] \in I_4^n, \\ b_{1\ell rst} := 0 \text{ and } b_{2\ell rst} := 0 & \text{otherwise.} \end{cases}$$

Their U-centered versions $(B_{k\ell rs})_{k,\ell,r,s\in[\![n]\!]}$ for $k=1,2$ are

$$B_{k\ell rst} := \begin{cases} b_{k\ell rst} - \dfrac{1}{n-4}\displaystyle\sum_{i=1}^{n} b_{kirst} - \dfrac{1}{n-4}\sum_{j=1}^{n} b_{k\ell jst} + \dfrac{1}{(n-3)(n-4)}\sum_{i,j=1}^{n} b_{kijst} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } [\ell, r, s, t] \in I_4^n, \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Then,

$$\binom{n}{6}^{-1} \sum_{i_1 < \cdots < i_6} h_R\Big((\boldsymbol{y}_{1i_1}, \boldsymbol{y}_{2i_1}), \ldots, (\boldsymbol{y}_{1i_6}, \boldsymbol{y}_{2i_6})\Big) = \frac{1}{n(n-1)(n-2)(n-5)} \sum_{[\ell,r,s,t] \in I_4^n} B_{1\ell rst} B_{2\ell rst},$$

which clearly has $O(n^4)$ complexity. This completes the proof. □

Appendix C

APPENDIX OF CHAPTER 4

C.1 Proofs

Throughout the proofs below, all the claims regarding conditional expectations, conditional variances, and conditional covariances are in the almost sure sense.

C.1.1 Proof of Proposition 4.2.2 (ξ_n^*)

Proof of Proposition 4.2.2 (ξ^*). Equation (21) in Dette et al. (2013) states that

$$\widehat{B}_{2n} - \widetilde{B}_{2n} - C_{1n} - C_{2n} = o_{\mathrm{P}}(n^{-1/2}),$$

but tracking a glitch in signs the equation should in fact be

$$\widehat{B}_{2n} - \widetilde{B}_{2n} + C_{1n} + C_{2n} = o_{\mathrm{P}}(n^{-1/2}).$$

Accordingly, a revised version of Equations (24)–(26) in Dette et al. (2013) shows that,

$$n^{1/2}(\xi_n^* - \xi) = \frac{12}{n^{1/2}} \sum_{i=1}^{n} (Z_i - \mathrm{E} Z_i) + o_{\mathrm{P}}(1) \tag{C.1.1}$$

where $Z_i := Z_{i,1} - Z_{i,2} - Z_{i,3}$ with

$$\begin{aligned}
Z_{i,1} &:= \int_0^1 \mathbb{1}\Big\{F_{X_2}\Big(X_{2i}\Big) \le u_2\Big\} \tau\Big(F_{X_1}\Big(X_{1i}\Big), u_2\Big) \mathrm{d}u_2,\\
Z_{i,2} &:= \int_0^1 \int_0^1 \mathbb{1}\Big\{F_{X_1}\Big(X_{1i}\Big) \le u_1\Big\} \tau\big(u_1, u_2\big) \frac{\partial}{\partial u_1} \tau\big(u_1, u_2\big) \mathrm{d}u_1 \mathrm{d}u_2,\\
Z_{i,3} &:= \int_0^1 \int_0^1 \mathbb{1}\Big\{F_{X_2}\Big(X_{2i}\Big) \le u_2\Big\} \tau\big(u_1, u_2\big) \frac{\partial}{\partial u_2} \tau\big(u_1, u_2\big) \mathrm{d}u_1 \mathrm{d}u_2,
\end{aligned}$$

$\tau(u_1, u_2) = \partial C(u_1, u_2)/\partial u_1$, and $C(u_1, u_2)$ is the copula of (X_1, X_2). Since the first term on the right hand side of (C.1.1) has finite variance (see computation on pages 34–35 of Dette et al. (2013)), we deduce that

$$\xi_n^* \xrightarrow{\mathsf{p}} \xi.$$

This completes the proof. □

C.1.2 Proof of Proposition 4.2.4(ii)

Proof of Proposition 4.2.4(ii). Applying (C.1.1), it holds under the null that

$$C(u_1, u_2) = u_1 u_2, \quad \tau(u_1, u_2) = u_2.$$

Accordingly,

$$Z_{i,1} = Z_{i,3} = \int_0^1 \mathbb{1}\Big\{F_{X_2}\Big(X_{2i}\Big) \le u_2\Big\} u_2 \mathrm{d}u_2 = \frac{1}{2}\Big[1 - \Big\{F_{X_2}\Big(X_{2i}\Big)\Big\}^2\Big] \quad \text{and} \quad Z_{i,2} = 0,$$

which yields

$$n^{1/2}\xi_n^* \xrightarrow{\mathsf{p}} 0. \tag{C.1.2}$$

This completes the proof. □

C.1.3 Proof of Remark 4.3.1

Proof of Remark 4.3.1. Recall that $f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta)$ denotes the density of $\boldsymbol{X}$ with Δ. Denote

$$\dot{\ell}(\boldsymbol{x}; \Delta) := \frac{\partial}{\partial \Delta} \log f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta).$$

These definitions make sense by Assumption 4.3.1(i),(ii), and we may write $\mathcal{I}_{\boldsymbol{X}}(0) = \mathrm{E}[\{\dot{\ell}(\boldsymbol{Y}; 0)\}^2]$. Notice that $\boldsymbol{Y}$ is distributed as $\boldsymbol{X}$ with $\Delta = 0$. Since $\boldsymbol{Y} = \mathbf{A}_\Delta^{-1}\boldsymbol{X}$ is an invertible linear transformation, the density of $\boldsymbol{X}$ can be expressed as

$$f_{\boldsymbol{X}}(\boldsymbol{x};\Delta) = |\det(\mathbf{A}_\Delta)|^{-1} f_{\boldsymbol{Y}}(\mathbf{A}_\Delta^{-1}\boldsymbol{x}),$$

where $f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{Y}}(y_1, y_2) = f_1(y_1)f_2(y_2)$. Direct computation yields

$$\dot{\ell}(\boldsymbol{x};0) = -x_1\Big\{\rho_2\Big(x_2\Big)\Big\} - x_2\Big\{\rho_1\Big(x_1\Big)\Big\}. \tag{C.1.3}$$

Thus $\mathrm{E}\{(Y_k)^2\} < \infty$ and $\mathrm{E}[\{\rho_k(Y_k)\}^2] < \infty$ for $k = 1, 2$ will imply $\mathcal{I}_{\boldsymbol{X}}(0) = \mathrm{E}[\{\dot{\ell}(\boldsymbol{Y};0)\}^2] < \infty$ under the Konijn alternatives. Also, $\mathrm{E}[\{\rho_k(Y_k)\}^2] < \infty$ implies that $\mathrm{E}\{\rho_k(Y_k)\} = 0$ by Lemma A.1 (Part A) in Johnson and Barron (2004). □

C.1.4 Proof of Example 4.3.1

Proof of Example 4.3.1. Assumption 4.3.1(i) is satisfied since $f_k(z) > 0$, $k = 1, 2$ for all real z. Assumption 4.3.1(iii) holds in view of (C.1.3); notice that $\dot{\ell}(\boldsymbol{x};0)$ can never always be 0. For Assumption 4.3.1(ii), if $\rho_k(z)$ is constant, then $f_k(z)$ is either constant or proportional to e^{Cz} with some constant C for all real z, which is impossible. Then Assumption 4.3.1 is satisfied.

Regarding the special case, without loss of generality, we can assume Y_1 and Y_2 to be standard normal or standard t-distributed. For the standard normal, we have $\rho_k(z) = -t$ and thus (4.3.2) is satisfied. For the standard t-distribution with ν_k degrees of freedom, we have $\rho_k(z) = -z(1 + 1/\nu_k)/(1 + z^2/\nu_k)$. It is easy to check (4.3.2) is satisfied when $\nu_k > 2$. □

C.1.5 Proof of Remark 4.3.2

Proof of Remark 4.3.2. Let $f_{\boldsymbol{X}}(\boldsymbol{x};\Delta)$ denote the density of $\boldsymbol{X}$ with Δ. Denote

$$\dot{\ell}(\boldsymbol{x};\Delta) := \frac{\partial}{\partial\Delta}\log f_{\boldsymbol{X}}(\boldsymbol{x};\Delta),$$

then we can write $\mathcal{I}_{\boldsymbol{X}}(0) = \mathrm{E}[\{\dot{\ell}(\boldsymbol{Y};0)\}^2]$, where $\boldsymbol{Y}$ is distributed as $\boldsymbol{X}$ with $\Delta = 0$. Direct computation yields

$$\dot{\ell}(\boldsymbol{x};0) = \frac{g(\boldsymbol{x}) - f_0(\boldsymbol{x})}{f_0(\boldsymbol{x})},$$

and thus

$$\begin{aligned}\mathcal{I}_{\boldsymbol{X}}(0) &= \mathrm{E}[\{\dot{\ell}(\boldsymbol{Y};0)\}^2] = \mathrm{E}[\{g(\boldsymbol{Y})/f_0(\boldsymbol{Y}) - 1\}^2] \\ &= \mathrm{E}[\{s(\boldsymbol{Y})\}^2] = \chi^2(G, F_0) := \int (\mathrm{d}G/\mathrm{d}F_0 - 1)^2 \mathrm{d}F_0.\end{aligned}$$

Since $s(\boldsymbol{x}) = g(\boldsymbol{x})/f_0(\boldsymbol{x}) - 1$ is continuous and both g and f_0 have compact support, $s(\boldsymbol{x})$ is bounded. Hence $\mathcal{I}_{\boldsymbol{X}}(0) < \infty$. □

C.1.6 Proof of Example 4.3.2

Proof of Example 4.3.2. To verify Assumption 4.3.2 for the Farlie alternatives, we first prove that G is a bonafide joint distribution function. The corresponding density g is given by

$$g(x_1, x_2) = f_1(x_1) f_2(x_2)[1 + \{1 - 2F_1(x_1)\}\{1 - 2F_2(x_2)\}],$$

which is a bonafide joint density function (Kössler and Rödel, 2007, Sec. 1.1.5). Then we have

$$s(\boldsymbol{x}) = g(\boldsymbol{x})/f_0(\boldsymbol{x}) - 1 = \{1 - 2F_1(x_1)\}\{1 - 2F_2(x_2)\}$$

and find that

$$\begin{aligned}&\mathrm{E}[s(\boldsymbol{Y})|Y_1] = \{1 - 2F_1(Y_1)\} \times \mathrm{E}\{1 - 2F_2(Y_2)\} = 0 \\ \text{and} \quad &\mathrm{E}[s(\boldsymbol{Y})|Y_2] = \mathrm{E}\{1 - 2F_1(Y_1)\} \times \{1 - 2F_2(Y_2)\} = 0.\end{aligned}$$

The proof is completed. □

C.1.7 Proof of Example 4.3.3

Proof of Example 4.3.3. We first verify that g is a bonafide joint density function. First since both h_1 and h_2 are bounded by 1,

$$|g(\boldsymbol{x})/f_0(\boldsymbol{x}) - 1| = |h_1(x_1)h_2(x_1)| \leq 1,$$

and thus $g(\boldsymbol{x}) \geq 0$. Then we write

$$g(x_1, x_2) = f_1(x_1)f_2(x_2) + f_1(x_1)h_1(x_1)f_2(x_2)h_2(x_2)$$

and

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 = &\int_{-\infty}^{\infty} f_1(x_1)\mathrm{d}x_1 \times \int_{-\infty}^{\infty} f_2(x_2)\mathrm{d}x_2 \\ &+ \int_{-\infty}^{\infty} f_1(x_1)h_1(x_1)\mathrm{d}x_1 \times \int_{-\infty}^{\infty} f_2(x_2)h_2(x_2)\mathrm{d}x_2 = 1,\end{aligned}$$

where

$$\int_{-\infty}^{\infty} f_1(x_1)h_1(x_1)\mathrm{d}x_1 < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} f_2(x_2)h_2(x_2)\mathrm{d}x_2 = 0$$

since $h_1(x_1), h_2(x_2)$ are bounded by 1 and $f_2(x_2)h_2(x_2) = -f_2(-x_2)h_2(-x_2)$. We also have

$$\begin{aligned}&\mathrm{E}[s(\boldsymbol{Y})|Y_1] = h_1(Y_1) \times \mathrm{E}[h_2(Y_2)] = h_1(Y_1)\int_{-\infty}^{\infty} f_2(x_2)h_2(x_2)\mathrm{d}x_2 = 0, \\ \text{and} \quad &\mathrm{E}[s(\boldsymbol{Y})|Y_2] = \mathrm{E}[h_1(Y_1)] \times h_2(Y_2) \quad \text{with } \mathrm{E}[h_1(Y_1)] = \int_{-\infty}^{\infty} f_1(x_1)h_1(x_1)\mathrm{d}x_1 \neq 0.\end{aligned}$$

The proof is completed. □

C.1.8 Proof of Theorem 4.3.1(i)

Proof of Theorem 4.3.1(i). **(A)** *This proof uses all of Assumption 4.3.1.* Let $\boldsymbol{Y}_i = (Y_{1i}, Y_{2i})$, $i = 1, \ldots, n$ be independent copies of $\boldsymbol{Y}$. Recall that $f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta)$ is the density of $\boldsymbol{X}$ with Δ.

Denote

$$L(\boldsymbol{x};\Delta) := \frac{f_{\boldsymbol{X}}(\boldsymbol{x};\Delta)}{f_{\boldsymbol{X}}(\boldsymbol{x};0)}, \quad \dot{\ell}(\boldsymbol{x};\Delta) := \frac{\partial}{\partial\Delta}\log f_{\boldsymbol{X}}(\boldsymbol{x};\Delta),$$

and define $\Lambda_n := \sum_{i=1}^n \log L(\boldsymbol{Y}_i;\Delta_n)$ and $T_n := \Delta_n \sum_{i=1}^n \dot{\ell}(\boldsymbol{Y}_i;0)$. These definitions make sense by Assumption 4.3.1(i),(ii).

To employ a corollary to Le Cam's third lemma, we wish to derive the joint limiting null distribution of $(-n^{1/2}\xi_n/3, \Lambda_n)$. Under the null hypothesis, it holds that $Y_{2[1]},\ldots,Y_{2[n]}$ are still independent and identically distributed, where $[i]$ is such that $Y_{1[1]} < \cdots < Y_{1[n]}$. In view of Angus (1995, Equation (9)), we have that under the null,

$$-\frac{1}{3}n^{1/2}\xi_n = n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]} + o_{\mathrm{P}}(1), \tag{C.1.4}$$

where

$$\begin{aligned}\Xi_{[i]} := \Big|F_{Y_2}\Big(Y_{2[i+1]}\Big) - F_{Y_2}\Big(Y_{2[i]}\Big)\Big| + F_{Y_2}\Big(Y_{2[i+1]}\Big)\Big\{1 - F_{Y_2}\Big(Y_{2[i+1]}\Big)\Big\} \\ + F_{Y_2}\Big(Y_{2[i]}\Big)\Big\{1 - F_{Y_2}\Big(Y_{2[i]}\Big)\Big\} - \frac{2}{3},\end{aligned} \tag{C.1.5}$$

and F_{Y_2} is the cumulative distribution function for Y_2. One readily verifies $|\Xi_{[i]}| \le 1$.

Using (C.1.4), the limiting null distribution of $(-n^{1/2}\xi_n/3, \Lambda_n)$ will be the same as that of $(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, \Lambda_n)$. To find the limiting null distribution of $(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, \Lambda_n)$, using the idea from Hájek and Šidák (1967, p. 210–214), we first find the limiting null distribution of

$$\begin{aligned}\Big(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, T_n\Big) &= \Big(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, n^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}(\boldsymbol{Y}_i;0)\Big) \\ &= \Big(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, n^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}(\boldsymbol{Y}_{[i]};0)\Big),\end{aligned}$$

where $\boldsymbol{Y}_{[i]} = (Y_{1[i]}, Y_{2[i]})$. To employ the Cramér–Wold device, we aim to show that under the null, for any real numbers a and b,

$$an^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]} + bn^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}(\boldsymbol{Y}_{[i]};0) \rightsquigarrow N\Big(0, 2a^2/45 + b^2\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)\Big). \tag{C.1.6}$$

The idea of the proof is to first show a conditional central limit result

$$an^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]} + bn^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}(\boldsymbol{Y}_{[i]};0)\Big|Y_{11},\ldots,Y_{1n} \rightsquigarrow N\Big(0, 2a^2/45 + b^2\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)\Big)$$

$$\text{for almost every sequence } Y_{11},\ldots,Y_{1n},\ldots, \tag{C.1.7}$$

and secondly deduce the desired unconditional central limit result.

To prove (C.1.7), we follow the idea put forward in the proof of Lemma 2.9.5 in van der Vaart and Wellner (1996). According to the central limit theorem for 1-dependent random variables (see, e.g., the Corollary in Orey, 1958, p. 546), the statement (C.1.7) is true if the following conditions hold: for almost every sequence $Y_{11},\ldots,Y_{1n},\ldots,$

$$\mathrm{E}_2\Big(W_{[i]}\Big) = 0, \tag{C.1.8}$$

$$\frac{1}{n}\mathrm{E}_2\Big\{\Big(\sum_{i=1}^{n}W_{[i]}\Big)^2\Big\} \to 2a^2/45 + b^2\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0), \tag{C.1.9}$$

$$\sum_{i=1}^{n}\mathrm{E}_2\Big(W_{[i]}^2\Big)\Big/\mathrm{E}_2\Big\{\Big(\sum_{i=1}^{n}W_{[i]}\Big)^2\Big\} \quad \text{is bounded}, \tag{C.1.10}$$

$$\text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_2\Big\{W_{[i]}^2 \times \mathbb{1}\Big(n^{-1/2}\Big|W_{[i]}\Big| > \epsilon\Big)\Big\} \to 0 \quad \text{for every } \epsilon > 0, \tag{C.1.11}$$

where E_2 denotes the expectation conditionally on $Y_{11},\ldots,Y_{1n}$, and

$$W_{[i]} := a\Xi_{[i]} + b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big) \quad \text{for } i = 1,\ldots,n-1,$$

$$\text{and} \quad W_{[n]} := b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[n]};0\Big). \tag{C.1.12}$$

We verify conditions (C.1.8)–(C.1.11) as follows, starting from (C.1.8). Under the null hypothesis, conditionally on $Y_{11}, \ldots, Y_{1n}$, we have that $Y_{2[1]}, \ldots, Y_{2[n]}$ are still independent and identically distributed as Y_2, which implies that $\mathrm{E}_2(\Xi_{[i]}) = 0$. We also deduce, by (C.1.3) and Assumption 4.3.1(ii), that

$$\mathrm{E}\Big\{\dot{\ell}\Big(\boldsymbol{Y}; 0\Big)\Big|Y_1\Big\} = 0, \tag{C.1.13}$$

and thus $\mathrm{E}_2\{\dot{\ell}(\boldsymbol{Y}_{[i]}; 0)\} = 0$. Then (C.1.8) follows by noticing that

$$\mathrm{E}_2(\Xi_{[i]}) = 0 \quad \text{and} \quad \mathrm{E}_2\{\dot{\ell}(\boldsymbol{Y}_{[i]}; 0)\} = 0. \tag{C.1.14}$$

For (C.1.9) and (C.1.10), we first claim that

$$\mathrm{Cov}_2\Big\{n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, n^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}(\boldsymbol{Y}_{[i]}; 0)\Big)\Big\} = 0, \tag{C.1.15}$$

where Cov_2 denotes the covariance conditionally on $Y_{11}, \ldots, Y_{1n}$. Recall that, under the null hypothesis, $Y_{2[1]}, \ldots, Y_{2[n]}$ are still independent and identically distributed as Y_2, conditionally on $Y_{11}, \ldots, Y_{1n}$. We obtain

$$\begin{aligned} &\mathrm{Cov}_2\Big\{\Big|F_{Y_2}\Big(Y_{2[i+1]}\Big) - F_{Y_2}\Big(Y_{2[i]}\Big)\Big|, \dot{\ell}\Big(\boldsymbol{Y}_{[i+1]}; 0\Big)\Big\} \\ &= \mathrm{Cov}_2\Big[\frac{1}{2}\Big\{F_{Y_2}\Big(Y_{2[i+1]}\Big)\Big\}^2 + \frac{1}{2}\Big\{1 - F_{Y_2}\Big(Y_{2[i+1]}\Big)\Big\}^2, \dot{\ell}\Big(\boldsymbol{Y}_{[i+1]}; 0\Big)\Big] \end{aligned} \tag{C.1.16}$$

by taking expectation with respect to $Y_{2[i]}$,

$$\begin{aligned} &\mathrm{Cov}_2\Big\{\Big|F_{Y_2}\Big(Y_{2[i+1]}\Big) - F_{Y_2}\Big(Y_{2[i]}\Big)\Big|, \dot{\ell}\Big(\boldsymbol{Y}_{[i]}; 0\Big)\Big\} \\ &= \mathrm{Cov}_2\Big[\frac{1}{2}\Big\{F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2 + \frac{1}{2}\Big\{1 - F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2, \dot{\ell}\Big(\boldsymbol{Y}_{[i]}; 0\Big)\Big] \end{aligned} \tag{C.1.17}$$

by taking expectation with respect to $Y_{2[i+1]}$, and

$$\mathrm{Cov}_2\Big\{\Big|F_{Y_2}\Big(Y_{2[i+1]}\Big) - F_{Y_2}\Big(Y_{2[i]}\Big)\Big|, \dot{\ell}\Big(\boldsymbol{Y}_{[j]}; 0\Big)\Big\} = 0 \quad \text{for all } j \neq i,\ i+1, \tag{C.1.18}$$

since $Y_{2[i]}, Y_{2[i+1]}$ are independent of $Y_{2[j]}$ with $j \neq i,\ i+1$, conditionally on $Y_{11}, \ldots, Y_{1n}$.

Taking into account (C.1.16)–(C.1.18), it follows that

$$\begin{aligned}
&\mathrm{Cov}_2\Big\{n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, n^{-1/2}\Delta_0\sum_{i=1}^{n}\dot{\ell}\Big(\boldsymbol{Y}_{[i]}\Big)\Big\} \\
&= n^{-1}\Delta_0\Big(\sum_{i=2}^{n}\mathrm{Cov}_2\Big[\frac{1}{2}\Big\{F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2+\frac{1}{2}\Big\{1-F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2,\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big] \\
&\qquad +\sum_{i=1}^{n-1}\mathrm{Cov}_2\Big[\frac{1}{2}\Big\{F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2+\frac{1}{2}\Big\{1-F_{Y_2}\Big(Y_{2[i]}\Big)\Big\}^2,\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big] \\
&\qquad +\sum_{i=2}^{n}\mathrm{Cov}_2\Big[F_{Y_2}\Big(Y_{2[i]}\Big)\Big\{1-F_{Y_2}\Big(Y_{2[i]}\Big)\Big\},\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big] \\
&\qquad +\sum_{i=1}^{n-1}\mathrm{Cov}_2\Big[F_{Y_2}\Big(Y_{2[i]}\Big)\Big\{1-F_{Y_2}\Big(Y_{2[i]}\Big)\Big\},\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big]\Big) \\
&= n^{-1}\Big[\sum_{i=2}^{n}\mathrm{Cov}_2\Big\{\frac{1}{2},\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}+\sum_{i=1}^{n-1}\mathrm{Cov}_2\Big\{\frac{1}{2},\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}\Big] \\
&= n^{-1}\Big[-\mathrm{Cov}_2\Big\{\frac{1}{2},\dot{\ell}\Big(\boldsymbol{Y}_{[1]};0\Big)\Big\}-\mathrm{Cov}_2\Big\{\frac{1}{2},\dot{\ell}\Big(\boldsymbol{Y}_{[n]};0\Big)\Big\}\Big]=0, \qquad \text{(C.1.19)}
\end{aligned}$$

where we notice that

$$\mathrm{E}\Big\{n^{-1}\Big|\dot{\ell}\Big(\boldsymbol{Y}_{[j]};0\Big)\Big|\Big\}\le \mathrm{E}\Big\{n^{-1}\sum_{i=1}^{n}\Big|\dot{\ell}\Big(\boldsymbol{Y}_i;0\Big)\Big|\Big\}=\mathrm{E}\Big|\dot{\ell}\Big(\boldsymbol{Y};0\Big)\Big|<\infty, \qquad \text{(C.1.20)}$$

for any given j. Then using (C.1.14)–(C.1.15) we can prove (C.1.9) as follows:

$$\begin{aligned}
\frac{1}{n}\mathrm{E}_2\Big\{\Big(\sum_{i=1}^{n}W_{[i]}\Big)^2\Big\} &= \frac{1}{n}\mathrm{E}_2\Big[\Big\{a\sum_{i=1}^{n-1}\Xi_{[i]}+b\Delta_0\sum_{i=1}^{n}\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}^2\Big] \\
&= \frac{1}{n}\mathrm{E}_2\Big[\Big(a\sum_{i=1}^{n-1}\Xi_{[i]}\Big)^2+\Big\{b\Delta_0\sum_{i=1}^{n}\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}^2\Big] \\
&= \frac{1}{n}\mathrm{E}_2\Big[\Big(a\sum_{i=1}^{n-1}\Xi_{[i]}\Big)^2+\Big\{b\Delta_0\sum_{i=1}^{n}\dot{\ell}\Big(\boldsymbol{Y}_{i};0\Big)\Big\}^2\Big] \\
&= \frac{2a^2(n-1)}{45n}+\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_2\Big[\Big\{b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{i};0\Big)\Big\}^2\Big]
\end{aligned}$$

$$\to 2a^2/45 + b^2\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0), \tag{C.1.21}$$

where the last step holds for almost all sequences $Y_{11}, \ldots, Y_{1n}, \ldots$ by the law of large numbers.

To verify (C.1.10), recalling (C.1.5) and using (C.1.14),(C.1.17), we obtain

$$\mathrm{E}_2\Big\{\Xi_{[i]} \times \Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\} = \mathrm{Cov}_2\Big\{\Xi_{[i]}, \Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\} = 0,$$

and moreover,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_2\Big(W_{[i]}^2\Big) &= \frac{1}{n}\bigg(\sum_{i=1}^{n-1}\mathrm{E}_2\Big[\Big\{a\Xi_{[i]} + b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}^2\Big] + \mathrm{E}_2\Big[\Big\{b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[n]};0\Big)\Big\}^2\Big]\bigg)\\ &= \frac{1}{n}\bigg(\sum_{i=1}^{n-1}\mathrm{E}_2\Big\{\Big(a\Xi_{[i]}\Big)^2\Big\} + \sum_{i=1}^{n}\mathrm{E}_2\Big[\Big\{b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big)\Big\}^2\Big]\bigg)\\ &= \frac{1}{n}\bigg(\frac{2a^2(n-1)}{45} + \sum_{i=1}^{n}\mathrm{E}_2\Big[\Big\{b\Delta_0\dot{\ell}\Big(\boldsymbol{Y}_i;0\Big)\Big\}^2\Big]\bigg).\end{aligned}$$

Hence we have, recalling (C.1.21),

$$\sum_{i=1}^{n}\mathrm{E}_2\Big(W_{[i]}^2\Big)\Big/\mathrm{E}_2\Big\{\Big(\sum_{i=1}^{n}W_{[i]}\Big)^2\Big\} = 1. \tag{C.1.22}$$

For proving (C.1.11), we recall that as given in (C.1.3)

$$\dot{\ell}\Big(\boldsymbol{Y}_{[i]};0\Big) = -Y_{1[i]}\Big\{\rho_2\Big(Y_{2[i]}\Big)\Big\} - Y_{2[i]}\Big\{\rho_1\Big(Y_{1[i]}\Big)\Big\}, \tag{C.1.23}$$

where $\rho_k(z) := f_k'(z)/f_k(z)$. The existence of finite second moments assumed in Assumption 4.3.1(iii), $\mathrm{E}\{(Y_1)^2\} < \infty$ and $\mathrm{E}[\{\rho_1(Y_1)\}^2] < \infty$, implies that

$$\max_{1\le i\le n} n^{-1/2}\Big|Y_{1i}\Big| \to 0 \quad \text{and} \quad \max_{1\le i\le n} n^{-1/2}\Big|\rho_1\Big(Y_{1i}\Big)\Big| \to 0 \tag{C.1.24}$$

for almost all sequences $Y_{11}, \ldots, Y_{1n}, \ldots$ (Barndorff-Nielsen, 1963, Theorem 5.2). Since

$|\Xi_{[i]}| \leq 1$, we have

$$\mathbb{1}\Big(n^{-1/2}\big|W_{[i]}\big| > \epsilon\Big) \leq \mathbb{1}\Big(|a|n^{-1/2} > \epsilon/3\Big) + \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{2[i]}\Big)\Big| > \epsilon/3\Big\}$$
$$+ \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|\rho_1\Big(Y_{1i}\Big)\Big|\Big) \times \Big|Y_{2[i]}\Big| > \epsilon/3\Big\}.$$

Then for every $\epsilon > 0$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_2\Big\{W_{[i]}^2 \times \mathbb{1}\Big(n^{-1/2}\Big|W_{[i]}\Big| > \epsilon\Big)\Big\} \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_2\Bigg(3\Big[a^2 + \Big\{Y_{1[i]} \times \rho_2\Big(Y_{2[i]}\Big)\Big\}^2 + \Big\{Y_{2[i]} \times \rho_1\Big(Y_{1[i]}\Big)\Big\}^2\Big] \\
&\qquad \times \Big[\mathbb{1}\Big(|a|n^{-1/2} > \epsilon/3\Big) + \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{2[i]}\Big)\Big| > \epsilon/3\Big\} \\
&\qquad\qquad + \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|\rho_1\Big(Y_{1i}\Big)\Big|\Big) \times \Big|Y_{2[i]}\Big| > \epsilon/3\Big\}\Big]\Bigg).
\end{aligned} \tag{C.1.25}$$

Here in (C.1.25) we have by (C.1.24) and dominated convergence theorem that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_2\Big[\mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{2[i]}\Big)\Big| > \epsilon/3\Big\}\Big] \\
&= \mathrm{E}_2\Big[\mathbb{1}\Big\{\Big|b\Big| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{21}\Big)\Big| > \epsilon/3\Big\}\Big] \to 0,
\end{aligned}$$

where

$$\mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{21}\Big)\Big| > \epsilon/3\Big\}\Big] \xrightarrow{\mathsf{p}} 0,$$

for almost all sequences $Y_{11}, \ldots, Y_{1n}, \ldots$. We also have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_2\Big[\Big\{Y_{1[i]}\rho_2\Big(Y_{2[i]}\Big)\Big\}^2 \times \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{2[i]}\Big)\Big| > \epsilon/3\Big\}\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} \Big(Y_{1[i]}\Big)^2 \mathrm{E}_2\Big[\Big\{\rho_2\Big(Y_{2[i]}\Big)\Big\}^2 \times \mathbb{1}\Big\{|b| \times \Big(\max_{1\leq i\leq n} n^{-1/2}\Big|Y_{1i}\Big|\Big) \times \Big|\rho_2\Big(Y_{2[i]}\Big)\Big| > \epsilon/3\Big\}\Big]
\end{aligned}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_{1[i]}\right)^2\right)\left(\mathrm{E}_2\left[\left\{\rho_2\left(Y_{21}\right)\right\}^2 \times \mathbb{1}\left\{|b| \times \left(\max_{1\le i\le n} n^{-1/2}\left|Y_{1i}\right|\right) \times \left|\rho_2\left(Y_{21}\right)\right| > \epsilon/3\right\}\right]\right)$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_{1i}\right)^2\right)\left(\mathrm{E}_2\left[\left\{\rho_2\left(Y_{21}\right)\right\}^2 \times \mathbb{1}\left\{|b| \times \left(\max_{1\le i\le n} n^{-1/2}\left|Y_{1i}\right|\right) \times \left|\rho_2\left(Y_{21}\right)\right| > \epsilon/3\right\}\right]\right)$$

$$\to 0,$$

where for almost all sequences $Y_{11}, \ldots, Y_{1n}, \ldots,$

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_{1i}\right)^2 \to \mathrm{E}\left\{\left(Y_1\right)^2\right\}$$

by the law of large numbers, and

$$\mathrm{E}_2\left[\left\{\rho_2\left(Y_{21}\right)\right\}^2 \times \mathbb{1}\left\{|b| \times \left(\max_{1\le i\le n} n^{-1/2}\left|Y_{1i}\right|\right) \times \left|\rho_2\left(Y_{21}\right)\right| > \epsilon/3\right\}\right] \to 0$$

by (C.1.24) and the dominated convergence theorem. We can deduce similar convergences for all the other summands in (C.1.25). Hence for almost all sequences $Y_{11}, \ldots, Y_{1n}, \ldots,$ all conditions (C.1.8)–(C.1.11) are satisfied. This completes the proof of (C.1.7). Moreover, the desired result (C.1.6) follows.

Finally, the Cramér–Wold device yields that under the null,

$$\left(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]}, T_n\right) \rightsquigarrow N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}2/45 & 0\\ 0 & \Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)\end{pmatrix}\right). \tag{C.1.26}$$

Furthermore, using ideas from Hájek and Šidák (1967, p. 210–214) (see also Gieser, 1993, Appx. B), we have under the null,

$$\Lambda_n - T_n + \Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)/2 \xrightarrow{\mathsf{p}} 0,$$

and thus under the null,

$$\Big(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]},\Lambda_n\Big)\rightsquigarrow N_2\bigg(\begin{pmatrix}0\\-\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)/2\end{pmatrix},\begin{pmatrix}2/45 & 0\\ 0 & \Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)\end{pmatrix}\bigg), \tag{C.1.27}$$

and $(-n^{1/2}\xi_n/3,\Lambda_n)$ has the same limiting null distribution by (C.1.4). Finally, we employ a corollary to Le Cam's third lemma (van der Vaart, 1998, Example 6.7) to obtain that, under the considered local alternative $H_{1,n}(\Delta_0)$ with any fixed $\Delta_0>0$, $-n^{1/2}\xi_n/3\rightsquigarrow N(0,2/45)$, and thus

$$n^{1/2}\xi_n\rightsquigarrow N(0,2/5). \tag{C.1.28}$$

This completes the proof for family (A).

(B) *This proof proceeds with only Assumption 4.3.2(i),(ii),(iv).* Let $\boldsymbol{Y}_i=(Y_{1i},Y_{2i})$, $i=1,\ldots,n$ be independent copies of $\boldsymbol{Y}$ (distributed as $\boldsymbol{X}$ with $\Delta=0$). Denote

$$L(\boldsymbol{x};\Delta):=\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\Delta)}{f_{\boldsymbol{X}}(\boldsymbol{x};0)},\quad \dot{\ell}(\boldsymbol{x};\Delta):=\frac{\partial}{\partial\Delta}\log f_{\boldsymbol{X}}(\boldsymbol{x};\Delta),$$

and define $\Lambda_n:=\sum_{i=1}^n\log L(\boldsymbol{Y}_i;\Delta_n)$ and $T_n:=\Delta_n\sum_{i=1}^n\dot{\ell}(\boldsymbol{Y}_i;0)$. Direct computation yields

$$L(\boldsymbol{x};\Delta)=\frac{(1-\Delta)f_0(\boldsymbol{x})+\Delta g(\boldsymbol{x})}{f_0(\boldsymbol{x})},\quad \dot{\ell}(\boldsymbol{x};0)=\frac{g(\boldsymbol{x})-f_0(\boldsymbol{x})}{f_0(\boldsymbol{x})},$$

and thus

$$\begin{aligned}\mathcal{I}_{\boldsymbol{X}}(0)&=\mathrm{E}[\{\dot{\ell}(\boldsymbol{Y};0)\}^2]=\mathrm{E}[\{g(\boldsymbol{Y})/f_0(\boldsymbol{Y})-1\}^2]\\&=\mathrm{E}[\{s(\boldsymbol{Y})\}^2]=\int(\mathrm{d}G/\mathrm{d}F_0-1)^2\mathrm{d}F_0.\end{aligned}$$

Similar to the proof for family (A), we proceed to determine the limiting null distribution of $(-n^{1/2}\xi_n/3,\Lambda_n)$. To this end, in view of the proof of Theorem 2 in Dhar et al. (2016), we first find the limiting null distribution of $(n^{-1/2}\sum_{i=1}^{n-1}\Xi_{[i]},T_n)$. The idea of deriving it is still to first show (C.1.7), then (C.1.6), and thus (C.1.26).

Next we verify conditions (C.1.8)–(C.1.11) for family (B). Notice that when we verify conditions (C.1.8)–(C.1.10) for family (A) (from (C.1.14) to (C.1.22)), we only use that

(1) under the null hypothesis, $Y_{2[1]}, \ldots, Y_{2[n]}$ are still independent and identically distributed as Y_2, conditionally on $Y_{11}, \ldots, Y_{1n}$,

(2) $\mathrm{E}\{\dot{\ell}(\boldsymbol{Y};0)|Y_1\} = 0$, and

(3) $0 < \mathcal{I}_{\boldsymbol{X}}(0) < \infty$.

The first property always holds under the null hypothesis. The latter two are assumed or implied in Assumption 4.3.2(ii) and Assumption 4.3.2(i),(iv), respectively. Hence we can verify conditions (C.1.8)–(C.1.10) for family (B) using the same arguments. The only difference lies in proving (C.1.11). Since $s(\boldsymbol{x}) = g(\boldsymbol{x})/f_0(\boldsymbol{x}) - 1$ is continuous and has compact support, it is bounded by some constant, say $C_s > 0$. We have by definition of $W_{[i]}$ in (C.1.12),

$$\left|W_{[i]}\right| \leq |a| + |b|\Delta_0 C_s,$$

and thus

$$\mathbb{1}\left(n^{-1/2}\left|W_{[i]}\right| > \epsilon\right) = 0 \quad \text{for all } n > \left(\frac{|a| + |b|\Delta_0 C_s}{\epsilon}\right)^2.$$

Then (C.1.11) follows by the dominated convergence theorem.

We have proven (C.1.26) for family (B). Furthermore, in the proof of Theorem 2 in Dhar et al. (2016), they showed that under the null,

$$\Lambda_n - T_n + \Delta_0^2 \mathcal{I}_{\boldsymbol{X}}(0)/2 \xrightarrow{\mathsf{p}} 0. \tag{C.1.29}$$

Thus under the null, we have (C.1.27) as well. The rest of the proof is to employ a corollary to Le Cam's third lemma (van der Vaart, 1998, Example 6.7) to obtain (C.1.28). □

C.1.9 Proof of Theorem 4.3.1(ii)

Proof of Theorem 4.3.1(ii). **(A)** *This proof uses all of Assumption 4.3.1.* Let $\boldsymbol{Y}_i = (Y_{1i}, Y_{2i})$ and $\boldsymbol{X}_i = (X_{1i}, X_{2i})$, $i = 1, \ldots, n$ be independent copies of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively. Here $\boldsymbol{X}$ depends on n with $\Delta = \Delta_n = n^{-1/2}\Delta_0$. Let $F^{(0)}$ and $F^{(a)}$ be the (joint) distribution functions of $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, respectively. Denote

$$L(\boldsymbol{x}; \Delta) := \frac{f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta)}{f_{\boldsymbol{X}}(\boldsymbol{x}; 0)}, \quad \dot{\ell}(\boldsymbol{x}; \Delta) := \frac{\partial}{\partial \Delta} \log f_{\boldsymbol{X}}(\boldsymbol{x}; \Delta),$$

and define $\Lambda_n := \sum_{i=1}^n \log L(\boldsymbol{Y}_i; \Delta_n)$ and $T_n := \Delta_n \sum_{i=1}^n \dot{\ell}(\boldsymbol{Y}_i; 0)$. These definitions make sense by Assumption 4.3.1(i),(ii).

In this proof we will consider the Hoeffding decomposition of μ_n under the null:

$$\mu_n = \sum_{\ell=1}^{m^\mu} \underbrace{\binom{n}{\ell}^{-1} \sum_{1 \le i_1 < \cdots < i_\ell \le n} \binom{m^\mu}{\ell} \widetilde{h}_\ell^\mu \Big\{ \Big(Y_{1i_1}, Y_{2i_1}\Big), \ldots, \Big(Y_{1i_\ell}, Y_{2i_\ell}\Big) \Big\}}_{H_{n,\ell}^\mu}, \tag{C.1.30}$$

where

$$\widetilde{h}_\ell^\mu(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\ell) := h_\ell^\mu(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\ell) - \mathrm{E}h^\mu - \sum_{k=1}^{\ell-1} \sum_{1 \le i_1 < \cdots < i_k \le \ell} \widetilde{h}_k^\mu(\boldsymbol{y}_{i_1}, \ldots, \boldsymbol{y}_{i_k}),$$

$$h_\ell^\mu(\boldsymbol{y}_1 \ldots, \boldsymbol{y}_\ell) := \mathrm{E}h^\mu(\boldsymbol{y}_1 \ldots, \boldsymbol{y}_\ell, \boldsymbol{Y}_{\ell+1}, \ldots, \boldsymbol{Y}_{m^\mu}), \quad \mathrm{E}h^\mu := \mathrm{E}h^\mu(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m^\mu}),$$

and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m^\mu}$ are m^μ independent copies of $\boldsymbol{Y}$. Here h^μ is the "symmetrized" kernel and m^μ is the order of the kernel function h^μ for $\mu \in \{D, R, \tau^*\}$ related to (4.2.5), (4.2.6), or (4.2.7):

$$\begin{aligned} h^D(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_5) := \frac{1}{5!} \sum_{1 \le i_1 \ne \cdots \ne i_5 \le 5} \frac{1}{4} & \\ \Big[\Big\{ \mathbb{1}\Big(y_{1i_1} \le y_{1i_5}\Big) - \mathbb{1}\Big(y_{1i_2} \le y_{1i_5}\Big) \Big\} \Big\{ \mathbb{1}\Big(y_{1i_3} \le y_{1i_5}\Big) - \mathbb{1}\Big(y_{1i_4} \le y_{1i_5}\Big) \Big\}\Big] & \\ \Big[\Big\{ \mathbb{1}\Big(y_{2i_1} \le y_{2i_5}\Big) - \mathbb{1}\Big(y_{2i_2} \le y_{2i_5}\Big) \Big\} \Big\{ \mathbb{1}\Big(y_{2i_3} \le y_{2i_5}\Big) - \mathbb{1}\Big(y_{2i_4} \le y_{2i_5}\Big) \Big\}\Big], & \end{aligned}$$

$$h^R(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_6) := \frac{1}{6!} \sum_{1 \le i_1 \ne \cdots \ne i_6 \le 6} \frac{1}{4}$$
$$\Big[\Big\{\mathbb{1}\Big(y_{1i_1} \le y_{1i_5}\Big) - \mathbb{1}\Big(y_{1i_2} \le y_{1i_5}\Big)\Big\}\Big\{\mathbb{1}\Big(y_{1i_3} \le y_{1i_5}\Big) - \mathbb{1}\Big(y_{1i_4} \le y_{1i_5}\Big)\Big\}\Big]$$
$$\Big[\Big\{\mathbb{1}\Big(y_{2i_1} \le y_{2i_6}\Big) - \mathbb{1}\Big(y_{2i_2} \le y_{2i_6}\Big)\Big\}\Big\{\mathbb{1}\Big(y_{2i_3} \le y_{2i_6}\Big) - \mathbb{1}\Big(y_{2i_4} \le y_{2i_6}\Big)\Big\}\Big],$$
$$h^{\tau^*}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_4) := \frac{1}{4!} \sum_{1 \le i_1 \ne \cdots \ne i_4 \le 4} \Big\{\mathbb{1}\Big(y_{1i_1}, y_{1i_3} < y_{1i_2}, y_{1i_4}\Big) + \mathbb{1}\Big(y_{1i_2}, y_{1i_4} < y_{1i_1}, y_{1i_3}\Big)$$
$$- \mathbb{1}\Big(y_{1i_1}, y_{1i_4} < y_{1i_2}, y_{1i_3}\Big) - \mathbb{1}\Big(y_{1i_2}, y_{1i_3} < y_{1i_1}, y_{1i_4}\Big)\Big\}$$
$$\Big\{\mathbb{1}\Big(y_{2i_1}, y_{2i_3} < y_{2i_2}, y_{2i_4}\Big) + \mathbb{1}\Big(y_{2i_2}, y_{2i_4} < y_{2i_1}, y_{2i_3}\Big)$$
$$- \mathbb{1}\Big(y_{2i_1}, y_{2i_4} < y_{2i_2}, y_{2i_3}\Big) - \mathbb{1}\Big(y_{2i_2}, y_{2i_3} < y_{2i_1}, y_{2i_4}\Big)\Big\},$$

and $m^D = 5$, $m^R = 6$, $m^{\tau^*} = 4$. We will omit the superscript μ in m^μ, h^μ, h^μ_ℓ, $\widetilde{h}^\mu_\ell$, and $H^\mu_{n,\ell}$ hereafter if no confusion is possible.

The proof is split into three steps. First, we prove that $F^{(a)}$ is contiguous to $F^{(0)}$ in order to employ Le Cam's third lemma (van der Vaart, 1998, Theorem 6.6). Next, we find the limiting null distribution of $(n\mu_n, \Lambda_n)$. Lastly, we employ Le Cam's third lemma to deduce the alternative distribution of $(n\mu_n, \Lambda_n)$.

Step I. In view of Gieser (1993, Sec. 3.2.1), Assumption 4.3.1 is sufficient for the contiguity: we have that $F^{(a)}$ is contiguous to $F^{(0)}$.

Step II. Next we need to derive the limiting distribution of $(n\mu_n, \Lambda_n)$ under null hypothesis. To this end, we first derive the limiting null distribution of $(nH_{n,2}, \Lambda_n)$, where $H_{n,2}$ is defined in (C.1.30). We write by the Fredholm theory of integral equations (Dunford and Schwartz, 1963, pages 1009, 1083, 1087) that

$$H_{n,2} = \frac{1}{n(n-1)} \sum_{i \ne j} \sum_{v=1}^{\infty} \lambda_v \psi_v\Big(Y_{1i}, Y_{2i}\Big)\psi_v\Big(Y_{1j}, Y_{2j}\Big),$$

where $\{\lambda_v, v = 1, 2, \ldots\}$ is an arrangement of $\{\lambda_{v_1,v_2}, v_1, v_2 = 1, 2, \ldots\}$, and ψ_v is the normalized eigenfunction associated with λ_v. For each positive integer K, define the "truncated"

U-statistic as

$$H_{n,2,K} := \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{v=1}^{K} \lambda_v \psi_v \Big(Y_{1i}, Y_{2i} \Big) \psi_v \Big(Y_{1j}, Y_{2j} \Big).$$

Notice that $nH_{n,2}$ and $nH_{n,2,K}$ can be written as

$$nH_{n,2} = \frac{n}{n-1} \Big(\sum_{v=1}^{\infty} \lambda_v \Big\{ n^{-1/2} \sum_{i=1}^{n} \psi_v \Big(Y_{1i}, Y_{2i} \Big) \Big\}^2 - \sum_{v=1}^{\infty} \lambda_v \Big[n^{-1} \sum_{i=1}^{n} \Big\{ \psi_v \Big(Y_{1i}, Y_{2i} \Big) \Big\}^2 \Big] \Big),$$

$$nH_{n,2,K} = \frac{n}{n-1} \Big(\sum_{v=1}^{K} \lambda_v \Big\{ n^{-1/2} \sum_{i=1}^{n} \psi_v \Big(Y_{1i}, Y_{2i} \Big) \Big\}^2 - \sum_{v=1}^{K} \lambda_v \Big[n^{-1} \sum_{i=1}^{n} \Big\{ \psi_v \Big(Y_{1i}, Y_{2i} \Big) \Big\}^2 \Big] \Big).$$

For a simpler presentation, let $S_{n,v}$ denote $n^{-1/2} \sum_{i=1}^{n} \psi_v(Y_{1i}, Y_{2i})$ hereafter.

To derive the limiting null distribution of $(nH_{n,2}, \Lambda_n)$, we first derive the limiting null distribution of $(nH_{n,2,K}, T_n)$ for each integer K. Observe that

$$\mathrm{E}(S_{n,v}) = 0, \quad \mathrm{Var}(S_{n,v}) = 1, \quad \mathrm{Cov}(S_{n,v}, T_n) \to d_v \Delta_0,$$
$$\mathrm{E}(T_n) = 0, \quad \mathrm{Var}(T_n) = \mathcal{I}_{\boldsymbol{X}}(0),$$

where $d_v := \mathrm{Cov}\{\psi_v(\boldsymbol{Y}), \dot{\ell}(\boldsymbol{Y}; 0)\}$ and $0 < \mathcal{I}_X(0) < \infty$ by Assumption 4.3.1. There exists at least one $v \geq 1$ such that $d_v \neq 0$. Indeed, applying Theorem 4.4 and Lemma 4.2 in Nandy et al. (2016) yields

$$\Big\{ \psi_v \Big(\boldsymbol{x} \Big), v = 1, 2, \ldots \Big\} = \Big\{ \psi_{1v_1} \Big(x_1 \Big) \psi_{2v_2} \Big(x_2 \Big), v_1, v_2 = 1, 2, \ldots \Big\},$$

where

$$\psi_{1v_1} \Big(x_1 \Big) \psi_{2v_2} \Big(x_2 \Big) := 2 \cos \Big\{ \pi v_1 F_{Y_1} \Big(x_1 \Big) \Big\} \cos \Big\{ \pi v_2 F_{Y_2} \Big(x_2 \Big) \Big\}$$

is associated with eigenvalue $\lambda^{\mu}_{v_1, v_2}$ defined in Proposition 4.2.4. Since

$$\mathrm{E} Y_k = \mathrm{E} \Big\{ \rho_{Y_k} \Big(Y_k \Big) \Big\} = 0,$$

$\{\psi_v(\boldsymbol{x}), v = 1, 2, \ldots\}$ forms a complete orthogonal basis for the family of functions of the form (C.1.3): $d_v = 0$ for all v thus entails

$$\mathcal{I}_X(0) = \mathrm{E}[\{L'(\boldsymbol{Y};0)\}^2] = \mathrm{E}\Big[\Big\{\sum_{v=1}^{\infty} d_v\psi_v\Big(Y_1, Y_2\Big)\Big\}^2\Big] = \sum_{v=1}^{\infty} d_v^2 = 0,$$

which contradicts Assumption 4.3.1(iii). Therefore, $d_{v^*} \neq 0$ for some v^*. Applying the multivariate central limit theorem (Bhattacharya and Ranga Rao, 1986, Equation (18.24)), we deduce that under the null,

$$(S_{n,1}, \ldots, S_{n,K}, T_n) \rightsquigarrow (\xi_1, \ldots, \xi_K, V_K),$$

where

$$(\xi_1, \ldots, \xi_K, V_K) \sim N_{K+1}\left(\begin{pmatrix} 0_K \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{I}_K & \Delta_0\boldsymbol{v} \\ \Delta_0\boldsymbol{v}^\top & \Delta_0^2\mathcal{I} \end{pmatrix}\right).$$

Here 0_K denotes a zero vector of dimension K, $\mathbf{I}_K$ denotes an identity matrix of dimension K, $\mathcal{I}$ is short for $\mathcal{I}_{\boldsymbol{X}}(0)$, and $\boldsymbol{v} = (d_1, \ldots, d_K)$. Thus V_K can be expressed as

$$\Big(\Delta_0^2\mathcal{I}\Big)^{1/2}\Big\{\sum_{v=1}^{K} c_v\xi_v + c_{0,K}\xi_0\Big\},$$

where $c_v := \mathcal{I}^{-1/2}d_v$, $c_{0,K} := (1 - \sum_{v=1}^{K} c_v^2)^{1/2}$, and ξ_0 is standard Gaussian and independent of $\xi_1, \ldots, \xi_K$. Then by the continuous mapping theorem (van der Vaart, 1998, Theorem 2.3) and Slutsky's theorem (van der Vaart, 1998, Theorem 2.8), we have under the null,

$$(nH_{n,2,K}, T_n) \rightsquigarrow \left(\sum_{v=1}^{K} \lambda_v\Big(\xi_v^2 - 1\Big),\ \Big(\Delta_0^2\mathcal{I}\Big)^{1/2}\Big(\sum_{v=1}^{K} c_v\xi_v + c_{0,K}\xi_0\Big)\right). \tag{C.1.31}$$

Moreover, we claim that under the null,

$$(nH_{n,2}, T_n) \rightsquigarrow \left(\sum_{v=1}^{\infty} \lambda_v\Big(\xi_v^2 - 1\Big),\ \Big(\Delta_0^2\mathcal{I}\Big)^{1/2}\Big(\sum_{v=1}^{\infty} c_v\xi_v + c_{0,\infty}\xi_0\Big)\right), \tag{C.1.32}$$

with $c_{0,\infty} := (1 - \sum_{v=1}^{\infty} c_v^2)^{1/2}$ via the following argument. Denote

$$M_K := \sum_{v=1}^{K} \lambda_v \Big(\xi_v^2 - 1\Big), \qquad V_K := \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(\sum_{v=1}^{K} c_v \xi_v + c_{0,K}\xi_0\Big),$$
$$M := \sum_{v=1}^{\infty} \lambda_v \Big(\xi_v^2 - 1\Big), \qquad \text{and} \quad V := \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(\sum_{v=1}^{\infty} c_v \xi_v + c_{0,\infty}\xi_0\Big).$$

To prove (C.1.32), it suffices to prove that for any real numbers a and b,

$$\Big|\mathrm{E}\Big\{\exp\Big(\mathsf{i}anH_{n,2} + \mathsf{i}bT_n\Big)\Big\} - \mathrm{E}\Big\{\exp\Big(\mathsf{i}aM + \mathsf{i}bV\Big)\Big\}\Big| \to 0 \quad \text{as } n \to \infty, \tag{C.1.33}$$

where i denotes the imaginary unit. We have

$$\begin{aligned}
&\Big|\mathrm{E}\Big\{\exp\Big(\mathsf{i}anH_{n,2} + \mathsf{i}bT_n\Big)\Big\} - \mathrm{E}\Big\{\exp\Big(\mathsf{i}aM + \mathsf{i}bV\Big)\Big\}\Big| \\
&\le \Big|\mathrm{E}\Big\{\exp\Big(\mathsf{i}anH_{n,2} + \mathsf{i}bT_n\Big)\Big\} - \mathrm{E}\Big\{\exp\Big(\mathsf{i}anH_{n,2,K} + \mathsf{i}bT_n\Big)\Big\}\Big| \\
&\quad + \Big|\mathrm{E}\Big\{\exp\Big(\mathsf{i}anH_{n,2,K} + \mathsf{i}bT_n\Big)\Big\} - \mathrm{E}\Big\{\exp\Big(\mathsf{i}aM_K + \mathsf{i}bV_K\Big)\Big\}\Big| \\
&\quad + \Big|\mathrm{E}\Big\{\exp\Big(\mathsf{i}aM_K + \mathsf{i}bV_K\Big)\Big\} - \mathrm{E}\Big\{\exp\Big(\mathsf{i}aM + \mathsf{i}bV\Big)\Big\}\Big| =: I + II + III, \quad \text{say,}
\end{aligned}$$

where in view of page 82 of Lee (1990) and Equation (4.3.10) in Koroljuk and Borovskich (1994),

$$I \le \mathrm{E}\Big|\exp\Big\{\mathsf{i}an\Big(H_{n,2} - H_{n,2,K}\Big)\Big\} - 1\Big| \le \Big\{\mathrm{E}\Big|an\Big(H_{n,2} - H_{n,2,K}\Big)\Big|^2\Big\}^{1/2} = \Big(\frac{2na^2}{n-1}\sum_{v=K+1}^{\infty} \lambda_v^2\Big)^{1/2}$$

and

$$\begin{aligned}
III &\le \mathrm{E}\Big|\exp\Big\{\mathsf{i}a\Big(M_K - M\Big) + \mathsf{i}b\Big(V_K - V\Big)\Big\} - 1\Big| \le \Big\{\mathrm{E}\Big|a\Big(M_K - M\Big) + b\Big(V_K - V\Big)\Big|^2\Big\}^{1/2} \\
&\le \Big\{2\Big(2a^2 \sum_{v=K+1}^{\infty} \lambda_v^2 + 2b^2\Delta_0^2\mathcal{I}\sum_{v=K+1}^{\infty} c_v^2\Big)\Big\}^{1/2}.
\end{aligned}$$

Since by Remark 3.1 in Nandy et al. (2016),

$$\sum_{v=1}^{\infty} \lambda_v^2 = \begin{cases} 1/8100 & \text{when } \mu = D, R, \\ 1/225 & \text{when } \mu = \tau^*, \end{cases} \quad \text{and} \quad \sum_{v=1}^{\infty} c_v^2 = \mathcal{I}^{-1} \sum_{v=1}^{\infty} d_v^2 = 1,$$

we conclude that, for any $\epsilon > 0$, there exists K_0 such that $I < \epsilon/3$ and $III < \epsilon/3$ for all n and all $K \geq K_0$. For this K_0, we have $II < \epsilon/3$ for all sufficiently large n by (C.1.31). These together prove (C.1.33). We also have, using the idea from Hájek and Šidák (1967, p. 210–214) (see also Gieser, 1993, Appendix B), that under the null

$$\Lambda_n - T_n + \Delta_0^2 \mathcal{I}/2 \xrightarrow{\mathsf{p}} 0. \tag{C.1.34}$$

Combining (C.1.32) and (C.1.34) yields that under the null,

$$(nH_{n,2}, \Lambda_n) \rightsquigarrow \Big(\sum_{v=1}^{\infty} \lambda_v \Big(\xi_v^2 - 1\Big), \ \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(\sum_{v=1}^{\infty} c_v \xi_v + c_{0,\infty} \xi_0 \Big) - \frac{\Delta_0^2 \mathcal{I}}{2} \Big). \tag{C.1.35}$$

Using the fact $H_{n,1} = 0$ and Equation (1.6.7) in Lee (1990, p. 30) yields that $(n\mu_n, \Lambda_n)$ has the same limiting distribution as (C.1.35) under the null.

Step III. Finally employing Le Cam's third lemma (van der Vaart, 1998, Theorem 6.6) we obtain that under the local alternative

$$\begin{aligned}
& \mathrm{P}\{n\mu_n \leq q_{1-\alpha} \mid H_{1,n}(\Delta_0)\} \\
\to\ & \mathrm{E}\Big[\mathbb{1}\Big\{ \sum_{v=1}^{\infty} \lambda_v \Big(\xi_v^2 - 1\Big) \leq q_{1-\alpha} \Big\} \times \exp\Big\{ \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(\sum_{v=1}^{\infty} c_v \xi_v + c_{0,\infty} \xi_0 \Big) - \frac{\Delta_0^2 \mathcal{I}}{2} \Big\}\Big] \\
\leq\ & \mathrm{E}\Big[\mathbb{1}\Big\{ \Big|\xi_{v^*}\Big| \leq \Big(\frac{q_{1-\alpha} + \sum_{v=1}^{\infty} \lambda_v}{\lambda_{v^*}} \Big)^{1/2} \Big\} \times \exp\Big\{ \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(\sum_{v=1}^{\infty} c_v \xi_v + c_{0,\infty} \xi_0 \Big) - \frac{\Delta_0^2 \mathcal{I}}{2} \Big\}\Big] \\
=\ & \mathrm{E}\Big[\mathbb{1}\Big\{ \Big|\xi_{v^*}\Big| \leq \Big(\frac{q_{1-\alpha} + \sum_{v=1}^{\infty} \lambda_v}{\lambda_{v^*}} \Big)^{1/2} \Big\} \times \exp\Big\{ \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big(c_{v^*} \xi_{v^*} + (1 - c_{v^*}^2)^{1/2} \xi_0 \Big) - \frac{\Delta_0^2 \mathcal{I}}{2} \Big\}\Big] \\
=\ & \Phi\Big\{ \Big(\frac{q_{1-\alpha} + \sum_{v=1}^{\infty} \lambda_v}{\lambda_{v^*}} \Big)^{1/2} - c_{v^*} \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big\} - \Phi\Big\{ - \Big(\frac{q_{1-\alpha} + \sum_{v=1}^{\infty} \lambda_v}{\lambda_{v^*}} \Big)^{1/2} - c_{v^*} \Big(\Delta_0^2 \mathcal{I}\Big)^{1/2} \Big\}
\end{aligned}$$

$$\leq 2\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\varphi\Big\{\Big|c_{v^*}\Big|\Big(\Delta_0^2\mathcal{I}\Big)^{1/2}-\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}\Big\}, \tag{C.1.36}$$

for some v^* such that $c_{v^*}=\mathcal{I}^{-1/2}d_{v^*}\neq 0$ and

$$\Delta_0 \geq \Big|c_{v^*}\Big|^{-1}\mathcal{I}^{-1/2}\Big(\frac{q_{1-\alpha}+\sum_{v=1}^{\infty}\lambda_v}{\lambda_{v^*}}\Big)^{1/2}, \tag{C.1.37}$$

where Φ and φ are the distribution function and density function of the standard normal distribution, respectively. Note that the right-hand side of (C.1.36) is monotonically decreasing as Δ_0 increases given (C.1.37). There exists a positive constant C_β such that (C.1.36) is smaller than $\beta/2$ as long as $\Delta_0\geq C_\beta$, regardless of whether c_{v^*} is positive or negative. This concludes the proof.

(B) *This proof uses Assumption 4.3.2(i),(iii),(iv).* Let $\boldsymbol{Y}_i=(Y_{1i},Y_{2i})$, $i=1,\dots,n$ be independent copies of $\boldsymbol{Y}$ (distributed as $\boldsymbol{X}$ with $\Delta=0$). Denote

$$L(\boldsymbol{x};\Delta):=\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\Delta)}{f_{\boldsymbol{X}}(\boldsymbol{x};0)},\quad \dot{\ell}(\boldsymbol{x};\Delta):=\frac{\partial}{\partial\Delta}\log f_{\boldsymbol{X}}(\boldsymbol{x};\Delta),$$

and define $\Lambda_n:=\sum_{i=1}^n\log L(\boldsymbol{Y}_i;\Delta_n)$ and $T_n:=\Delta_n\sum_{i=1}^n\dot{\ell}(\boldsymbol{Y}_i;0)$. Direct computation yields

$$L(\boldsymbol{x};\Delta)=\frac{(1-\Delta)f_0(\boldsymbol{x})+\Delta g(\boldsymbol{x})}{f_0(\boldsymbol{x})},\quad \dot{\ell}(\boldsymbol{x};0)=\frac{g(\boldsymbol{x})-f_0(\boldsymbol{x})}{f_0(\boldsymbol{x})},$$

and thus

$$\begin{aligned}\mathcal{I}_{\boldsymbol{X}}(0)&=\mathrm{E}[\{\dot{\ell}(\boldsymbol{Y};0)\}^2]=\mathrm{E}[\{g(\boldsymbol{Y})/f_0(\boldsymbol{Y})-1\}^2]\\&=\mathrm{E}[\{s(\boldsymbol{Y})\}^2]=\int(\mathrm{d}G/\mathrm{d}F_0-1)^2\mathrm{d}F_0.\end{aligned}$$

This is similar to the proof for family (A). The only difference lies in proving the existence of at least one $v\geq 1$ such that $d_v\neq 0$, where $d_v:=\mathrm{Cov}[\psi_v(\boldsymbol{Y}),\dot{\ell}(\boldsymbol{Y};0)]$. Now $\dot{\ell}(\boldsymbol{x};0)=s(\boldsymbol{x})$ is not of the form (C.1.3), and $\{\psi_v(\boldsymbol{x}),v=1,2,\dots\}$ does not necessarily form a complete orthogonal basis for the family of functions of $s(\boldsymbol{x})$. However, recall that

$$\Big\{\psi_v\Big(\boldsymbol{x}\Big), v = 1, 2, \dots\Big\} = \Big\{\psi_{1v_1}\Big(x_1\Big)\psi_{2v_2}\Big(x_2\Big), v_1, v_2 = 1, 2, \dots\Big\},$$

where

$$\psi_{1v_1}\Big(x_1\Big)\psi_{2v_2}\Big(x_2\Big) := 2\cos\Big\{\pi v_1 F_{Y_1}\Big(x_1\Big)\Big\}\cos\Big\{\pi v_2 F_{Y_2}\Big(x_2\Big)\Big\}.$$

Since

$$\Big\{\psi_{1v_1}\Big(x_1\Big)\psi_{2v_2}\Big(x_2\Big), v_1, v_2 = 0, 1, 2, \dots\Big\}$$

forms a complete orthogonal basis of the set of square integrable functions, $d_v = 0$ for all $v \geq 1$ thus entails $s(\boldsymbol{x}) = h_1(x_1) + h_2(x_2)$ for some functions h_1, h_2, where $h_k(x_k)$ depends only on x_k for $k = 1, 2$. This contradicts Assumption 4.3.2(iii). □

C.1.10 Proof of Proposition 4.3.1

Proof of Proposition 4.3.1. **(A)** *This proof uses all of Assumption 4.3.1.* Let $\boldsymbol{Y}_i = (Y_{1i}, Y_{2i})$ and $\boldsymbol{X}_i = (X_{1i}, X_{2i})$, $i = 1, \dots, n$ be independent copies of $\boldsymbol{Y}$ and $\boldsymbol{X}$ with $\Delta = \Delta_n = n^{-1/2}\Delta_0$, respectively. Let $F^{(0)}$ and $F^{(a)}$ be the (joint) distribution functions of $(\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n)$ and $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$, respectively, and let $F_i^{(0)}$ and $F_i^{(a)}$ be the distribution functions of $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$, respectively.

The total variation distance between two distribution functions G and F on the same real probability space is defined as

$$\mathrm{TV}(G, F) := \sup_A \Big|\mathrm{P}_G(A) - \mathrm{P}_F(A)\Big|,$$

where A is taken over the Borel field and $\mathrm{P}_G, \mathrm{P}_F$ are respective probability measures induced by G and F. Furthermore, if G is absolutely continuous with respect to F, the Hellinger distance between G and F is defined as

$$\mathrm{HL}(G, F) := \Big[\int 2\Big\{1 - (\mathrm{d}G/\mathrm{d}F)^{1/2}\Big\}\mathrm{d}F\Big]^{1/2}.$$

By Assumption 4.3.1(i), $\mathrm{HL}(F^{(a)}, F^{(0)})$ is well-defined. It suffices to prove that for any small $0 < \beta < 1-\alpha$, there exists $\Delta_0 = c_\beta$ such that, for all sufficiently large n, $\mathrm{TV}(F^{(a)}, F^{(0)}) < \beta$, which is implied by $\mathrm{HL}(F^{(a)}, F^{(0)}) < \beta$ using the relation (Tsybakov, 2009, Equation (2.20))

$$\mathrm{TV}\Big(F^{(a)}, F^{(0)}\Big) \le \mathrm{HL}\Big(F^{(a)}, F^{(0)}\Big).$$

We also know that (Tsybakov, 2009, page 83)

$$1 - \frac{1}{2}\mathrm{HL}^2\Big(F^{(a)}, F^{(0)}\Big) = \prod_{i=1}^{n}\Big\{1 - \frac{1}{2}\mathrm{HL}^2\Big(F_i^{(a)}, F_i^{(0)}\Big)\Big\}.$$

We then aim to evaluate $\mathrm{HL}^2(F^{(a)}, F^{(0)})$ in terms of $\mathcal{I}_{\boldsymbol{X}}(0)$ and Δ_0. By definition,

$$\frac{1}{2}\mathrm{HL}^2\Big(F_i^{(a)}, F_i^{(0)}\Big) = \mathrm{E}\Big[1 - \Big\{L\Big(\boldsymbol{Y}_i; \Delta_n\Big)\Big\}^{1/2}\Big].$$

Given Assumption 4.3.1, we deduce in view of Gieser (1993, Appendix B) that

$$n\mathrm{E}\Big[1 - \Big\{L\Big(\boldsymbol{Y}_i; \Delta_n\Big)\Big\}^{1/2}\Big] = \mathrm{E}\Big(\sum_{i=1}^{n}\Big[1 - \Big\{L\Big(\boldsymbol{Y}_i; \Delta_n\Big)\Big\}^{1/2}\Big]\Big) \to \frac{\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)}{8}.$$

Therefore,

$$1 - \frac{1}{2}\mathrm{HL}^2\Big(F^{(a)}, F^{(0)}\Big) \to \exp\Big\{-\frac{\Delta_0^2\mathcal{I}_{\boldsymbol{X}}(0)}{8}\Big\}.$$

The desired result follows by taking $c_\beta > 0$ such that

$$\exp\Big\{-\frac{c_\beta^2\mathcal{I}_{\boldsymbol{X}}(0)}{8}\Big\} = 1 - \frac{\beta^2}{8}.$$

(B) *This proof requires Assumption 4.3.2(i),(iv).* This is similar to the proof for family

(A), but here we will use the relation (Tsybakov, 2009, Equation (2.27))

$$\mathrm{TV}\Big(F^{(a)}, F^{(0)}\Big) \le \Big\{\chi^2\Big(F^{(a)}, F^{(0)}\Big)\Big\}^{1/2},$$

where the chi-square distance between two distribution functions G and F on the same real probability space such that G is absolutely continuous with respect to F is defined as

$$\chi^2(G, F) := \int \Big(\mathrm{d}G/\mathrm{d}F - 1\Big)^2 \mathrm{d}F.$$

Here $\chi^2(F^{(a)}, F^{(0)})$ is well-defined by Assumption 4.3.2(i). We also know that (Tsybakov, 2009, page 86)

$$1 + \chi^2\Big(F^{(a)}, F^{(0)}\Big) = \prod_{i=1}^{n} \Big\{1 + \chi^2\Big(F_i^{(a)}, F_i^{(0)}\Big)\Big\}.$$

Next we aim to evaluate $\chi^2(F^{(a)}, F^{(0)})$ in terms of $\mathcal{I}_{\boldsymbol{X}}(0) = \chi^2(G, F_0)$ and Δ_0. Here $0 < \mathcal{I}_{\boldsymbol{X}}(0) < \infty$ by Assumption 4.3.2(i),(iv). We have by definition that

$$\chi^2\Big(F_i^{(a)}, F_i^{(0)}\Big) = \chi^2\Big((1-\Delta_n)F_0 + \Delta_n G, F_0\Big) = \Delta_n^2 \chi^2(G, F_0) = n^{-1}\Delta_0^2 \chi^2(G, F_0).$$

Therefore, it holds that

$$1 + \chi^2\Big(F^{(a)}, F^{(0)}\Big) \to \exp\Big\{\Delta_0^2 \chi^2\Big(G, F_0\Big)\Big\}.$$

The desired result follows by taking $c_\beta > 0$ such that

$$\exp\Big\{c_\beta^2 \chi^2\Big(G, F_0\Big)\Big\} = 1 + \frac{\beta^2}{4}.$$

This completes the proof. $\square$