Bernoulli **18**(1), 2012, 229–251
DOI: 10.3150/10-BEJ326

Nonparametric multivariate rank tests and their unbiasedness

JANA JUREČKOVÁ[1] and JAN KALINA[2]

[1]*Department of Probability and Statistics, Charles University in Prague, Sokolovská 83, CZ-186 75 Prague 8, Czech Republic. E-mail: jurecko@karlin.mff.cuni.cz*
[2]*EUROMISE Center, Department of Medical Informatics, Institute of Computer Science of the Academy of Sciences of CR, v.v.i., Pod Vodárenskou věží 2, CZ-182 07 Prague 8, Czech Republic.*
E-mail: kalina@euromise.cz

Although unbiasedness is a basic property of a good test, many tests on vector parameters or scalar parameters against two-sided alternatives are not finite-sample unbiased. This was already noticed by Sugiura [*Ann. Inst. Statist. Math.* **17** (1965) 261–263]; he found an alternative against which the Wilcoxon test is not unbiased. The problem is even more serious in multivariate models. When testing the hypothesis against an alternative which fits well with the experiment, it should be verified whether the power of the test under this alternative cannot be smaller than the significance level. Surprisingly, this serious problem is not frequently considered in the literature.

The present paper considers the two-sample multivariate testing problem. We construct several rank tests which are finite-sample unbiased against a broad class of location/scale alternatives and are finite-sample distribution-free under the hypothesis and alternatives. Each of them is locally most powerful against a specific alternative of the Lehmann type. Their powers against some alternatives are numerically compared with each other and with other rank and classical tests. The question of affine invariance of two-sample multivariate tests is also discussed.

Keywords: affine invariance; contiguity; Kolmogorov–Smirnov test; Lehmann alternatives; Liu–Singh test; Psi test; Savage test; two-sample multivariate model; unbiasedness; Wilcoxon test

1. Introduction

1.1. Two-sample multivariate tests

A frequent practical problem is that we have two data clouds of p-dimensional observations with generally unknown distributions F and G, and we wish to test the hypothesis that they both come from the same distribution $F \equiv G$, continuous but unknown. Desirable properties of a test of such a hypothesis $\mathbf{H}$ are: (i) being distribution-free under $\mathbf{H}$; (ii) being affine invariant with respect to changes of coordinate system; (iii) being consistent against any fixed alternative; and (iv) being finite-sample unbiased against a broad class of alternatives of interest. Unfortunately, a test satisfying all these conditions does not exist in the multivariate setup.

Many authors have tried to attack this problem, emphasizing some of the above properties. Their ideas were often concentrated either on some geometric entities of the data clouds or on the affine invariance of the testing problem. Naturally, the ranks or the signed ranks of geometric entities of data are invariant under many transformations and provide a useful and simple tool

for testing. The papers extending methods based on ranks or other nonparametric methods to the multivariate setup use data depths, Oja medians, multivariate sign functions and other tools. In this context, we should mention the papers by Chaudhuri and Sengupta [5], Choi and Marden [6,7], Hallin and Pandaveine [13], Hetmansperger *et al.* [17], Liu [21,23], Liu and Singh [24], Oja and Randles [31], Oja *et al.* [30], Puri and Sen [33], Randles and Peters [34], Topchii *et al.* [40], Tukey [41], Zuo and He [43] and a recent excellent review by Oja [29].

Other authors have constructed various permutation tests: Bickel [3], Brown [4], Hall and Tajvidi [14], Neuhaus and Zhu [26], Oja [28], Wellner [42] and others. Tests based on distances between observations were considered by Baringhaus and Franz [2], Friedman and Rafsky [8], Henze [15], Maa *et al.* [25], Rosenbaum [35] and Schilling [37]; the latter also compared the simulated powers of his test with that of the Kolmogorov–Smirnov two-sample test.

The proposed tests were typically consistent against distant alternatives and some of them were affine invariant. The authors often derived the asymptotic null distributions of the test criteria and some derived the asymptotic powers under contiguous alternatives. Many authors illustrated the powers on the simulated data, often normal, and compared them with the power of the Hotelling T^2 test. However, only in exceptional cases did they check whether the test was unbiased against alternatives of interest.

1.2. Unbiased tests

Let Φ be a test of hypothesis $\{\mathbf{H}:$ distribution F of random vector $\mathbf{X}$ belongs to the set $\mathcal{H}\}$ against the alternative $\{\mathbf{K}:$ distribution of $\mathbf{X}$ belongs to the set $\mathcal{K}\}$. Consider the tests of size α, $0 < \alpha < 1$, where α is the chosen significance level, that is, the tests satisfying $\sup_{F\in\mathcal{H}} \mathbb{E}_F[\Phi(\mathbf{X})] \le \alpha$. The test Φ is unbiased if it satisfies

$$\sup_{F\in\mathcal{H}} \mathbb{E}_F[\Phi(\mathbf{X})] \le \alpha \quad \text{and} \quad \inf_{F\in\mathcal{K}} \mathbb{E}_F[\Phi(\mathbf{X})] \ge \alpha.$$

This is a natural property of a test; it means that the power of a test should not be smaller than the permitted error of the first kind. If the test rejects the hypothesis with a probability less than α under the alternative of interest, then we can hardly recommend the test to the experimenter. Note that if there exists a uniformly most powerful test, then it is always unbiased. If the optimal test of size α does not exist because the family of α-tests is too broad, then we should restrict ourselves to a pertinent subfamily of tests, and the family of unbiased tests of size α is the most natural subfamily. We refer to Lehmann's monograph [22] for an excellent account of unbiased tests.

Many tests criteria have asymptotic normal distributions under the hypothesis as well as under the local alternatives – these are *asymptotically* locally unbiased. However, the practice always works with a finite number of observations. The asymptotic distribution only approximates well the central part of the finite-sample distribution; elsewhere, it can stretch the truth and sometimes is only valid for a huge number of observations. To calculate the finite-sample power of a test is sometimes difficult; in any case, as a first step, we should be sure that the test is unbiased against the alternatives under consideration, at least locally in a neighborhood of the hypothesis.

Unfortunately, many authors have not specified the alternatives against which their tests are (locally) unbiased. The alternative is typically more important for an experimenter than the hypothesis because it describes his/her scientific conjecture. Some papers, for example, [18,19,38,39], show that the tests are not automatically finite-sample unbiased. While the univariate two-sample Wilcoxon test, for example, is always unbiased against one-sided alternatives, it is generally not unbiased against two-sided alternatives, even not with equal sample sizes (see [38,39]). The test is locally unbiased against two-sample alternatives only under some conditions on the hypothetical distribution of observations (e.g., when it is symmetric). Amrhein [1] demonstrated the same phenomenon for the one-sample Wilcoxon test. Hence, the finite-sample unbiasedness of some tests cited above, and others described in the literature, is still an open question.

To illustrate this problem more precisely, consider a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with distribution function $F(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$, and density $f(\mathbf{x}, \boldsymbol{\theta})$ (not necessarily Lebesgue), which has bounded third derivatives in components of $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$ and a positive definite Fisher information matrix. We wish to test $\mathbf{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\mathbf{K} : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ using the test Φ of size α, that is, $\mathbb{E}_{\theta_0}[\Phi(\mathbf{X})] = \alpha$. We then have the following expansion of the power function of Φ around $\boldsymbol{\theta}_0$ (see [19]):

$$\begin{aligned}\mathbb{E}_{\theta_0}\Phi(\mathbf{X}) &= \alpha + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbb{E}_{\theta_0}\left\{\Phi(\mathbf{X})\frac{(\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}_0))}{f(\mathbf{X}, \boldsymbol{\theta}_0)}\right\} \\ &\quad + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbb{E}_{\theta_0}\left\{\Phi(\mathbf{X})\frac{[\ddot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}_0)]}{f(\mathbf{X}, \boldsymbol{\theta}_0)}\right\}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathcal{O}(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^3),\end{aligned} \tag{1.1}$$

where

$$(\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\theta})) = \left(\frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_p}\right)^\top, \qquad [\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\theta})] = \left[\frac{\partial^2 f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j\, \partial \theta_k}\right]_{j,k=1}^p.$$

The test Φ is locally unbiased if the second term on the right-hand side of (1.1) is nonnegative. If θ is a scalar parameter and we consider the one-sided alternative $\mathbf{K} : \theta > \theta_0$, then there always exists an unbiased test. However, the alternative for a vector $\boldsymbol{\theta}$ is only two-sided and the local unbiasedness of Φ is guaranteed only when

$$\mathbb{E}_{\theta_0}\left\{\Phi(\mathbf{X})\frac{(\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}_0))}{f(\mathbf{X}, \boldsymbol{\theta}_0)}\right\} = \mathbf{0}. \tag{1.2}$$

However, (1.2) is generally true only for f satisfying special conditions, which cannot easily be verified for unknown f. If the test Φ does not satisfy (1.2), then the second term in (1.1) can be negative for some $\boldsymbol{\theta}$ and hence the power of Φ can be less than α. We refer to Grose and King [10], who imposed condition (1.2) when constructing a locally unbiased two-sided version of the Durbin–Watson test.

1.3. Outline of the paper

We shall propose three classes of multivariate two-sample tests, based on the ranks of suitable distances of multivariate observations. One test is based on the ranks of distances of observations

from the origin, while the others are based on the ranks of their interpoint distances. The natural alternatives state that either the distances of the second sample from the origin are stochastically larger than those of the first sample, or that the distances of the **Y**'s from the **X**'s are stochastically larger than those of the **X**'s from each other. The proposed tests are unbiased because our natural alternatives are one-sided (in the distances). Moreover, the proposed rank tests are distribution-free under the hypothesis as well as under alternatives of the Lehmann type, and they are consistent against general alternatives (properties (i), (iii) and (iv)). The distribution-free property is important because we need not determine the distribution of distances when performing the test.

The tests are described in Section 2, which starts with some invariance considerations (cf. desired property (ii) of the test). It is shown that the proposed tests based on the ranks of distances, as well as the Liu–Singh [24] tests based on the ranks of depths, are distribution-free against some monotone alternatives of the Lehmann type with respect to which they are finite-sample unbiased. Section 3 describes the contiguity of these alternatives with respect to the hypothesis, which enables us to derive the local asymptotic powers of the tests. The powers of tests are compared numerically under finite N, as well as asymptotically under $N \to \infty$. The proposed tests are also compared with the tests of Liu and Singh, and of [17], using a reference to numerical results of [43]. In Section 4 we compare the empirical powers of the proposed tests with those of the Hotelling test under the bivariate normal and bivariate Cauchy distributions. The contiguity of the Lehmann-type alternatives is proved in the Appendix.

2. Multivariate two-sample rank tests

2.1. Remarks to affine invariance

Consider two independent samples $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$ and $\mathcal{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ from two p-variate populations with continuous distribution functions $F^{(p)}$ and $G^{(p)}$, respectively, with respective means and dispersion matrices $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$. The problem is to test the hypothesis $\mathbf{H}_0 : F^{(p)} \equiv G^{(p)}$ (along with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$) against an alternative $\mathbf{H}_1$, where either $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \neq (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ or where $F^{(p)}$ and $G^{(p)}$ are not of the same functional form. We denote by $(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$ the pooled sample with $\mathbf{Z}_i = \mathbf{X}_i, i = 1, \ldots, m$, and $\mathbf{Z}_{m+j} = \mathbf{Y}_j, j = 1, \ldots, n, m + n = N$. The hypothesis and the alternative are invariant under affine transformations:

$$\mathcal{G} : \{\mathbf{Z} \to \mathbf{a} + \mathbf{B}\mathbf{Z}\} \text{ with } \mathbf{a} \in \mathbb{R}^p \text{and } \mathbf{B} \text{ a nonsingular } p \times p \text{ matrix.} \tag{2.1}$$

More precisely, the hypothesis and alternative remain true even after the transformation $g \in \mathcal{G}$ of the data, and we are looking for invariant tests whose criteria are invariant with respect to $g \in \mathcal{G}$. The invariant tests depend on the data only by means of a *maximal invariant* of $\mathcal{G}$ [22]. Obenchain [27] showed that the maximal invariant with respect to $\mathcal{G}$ is

$$\begin{gathered} \mathbf{T}(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = [(\mathbf{Z}_i - \bar{\mathbf{Z}}_N)^\top \mathbf{V}_N^{-1} (\mathbf{Z}_j - \bar{\mathbf{Z}}_N)]_{i,j=1}^N, \\ \text{where } \bar{\mathbf{Z}}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i, \mathbf{V}_N = \sum_{i=1}^N (\mathbf{Z}_i - \bar{\mathbf{Z}}_N)(\mathbf{Z}_i - \bar{\mathbf{Z}}_N)^\top. \end{gathered} \tag{2.2}$$

Then, $\mathbf{T}(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$ is the projection matrix associated with the space spanned by the columns of the matrix $[\mathbf{Z}_1 - \bar{\mathbf{Z}}_N, \ldots, \mathbf{Z}_N - \bar{\mathbf{Z}}_N]$. In particular, under $\mathbf{a} \equiv \mathbf{0}$, the maximal invariant of the group

$$\begin{aligned} &\mathcal{G}_0 : \{\mathbf{Z} \to \mathbf{BZ}\} \quad \text{is equal to} \quad \mathbf{T}_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = [\mathbf{Z}_i^\top (\mathbf{V}_N^0)^{-1} \mathbf{Z}_j]_{i,j=1}^N, \\ &\text{where } \mathbf{V}_N^0 = \sum_{i=1}^N \mathbf{Z}_i \mathbf{Z}_i^\top. \end{aligned} \tag{2.3}$$

Moreover, one of the maximal invariants with respect to the *group of shifts in location*

$$\mathcal{G}_1 : \mathbf{Z} \longrightarrow \mathbf{Z} + \mathbf{a}, \qquad \mathbf{a} \in \mathbb{R}^p, \tag{2.4}$$

is $\mathbf{T}_1(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = (\mathbf{Z}_2 - \mathbf{Z}_1, \ldots, \mathbf{Z}_N - \mathbf{Z}_1)$.

The well-known two-sample Hotelling T^2 test is based on the criterion

$$\mathcal{T}_{mn}^2 = (\bar{\mathbf{X}}_m - \bar{\mathbf{Y}}_n)^\top \mathbf{V}_N^{-1} (\bar{\mathbf{X}}_m - \bar{\mathbf{Y}}_n). \tag{2.5}$$

The test is invariant with respect to $\mathcal{G}$ and is optimal unbiased against two-sample normal alternatives with $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Its asymptotic null distribution, when both sample sizes m, n tend to infinity, does not depend on the normality. If $m, n \to \infty$ and $\frac{m}{n} \to 1$, then the asymptotic distribution of $\mathcal{T}_{mn}^2$ does not change even when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, but only in this case (see [22]). Its finite sample unbiasedness is not guaranteed under a nonnormal underlying distribution.

If we wish to construct a nonparametric two-sample test which is distribution-free and affine invariant with respect to $\mathcal{G}$, we expect it to be based on the ranks of some components of $\mathbf{T}$ in (2.2) or on the relevant empirical Mahalanobis distances of points $\mathbf{Z}_i, \mathbf{Z}_j$, $1 \le i, j \le N$. The ranks of distances are invariant with respect to continuous increasing functions of the distances; however, in our case, the data themselves are transformed, rather than their distances. The proper form of the rank test criterion based on the Mahalanobis distances and its unbiasedness against alternatives of interest is the subject of a forthcoming study. The rank tests considered in the present paper are easier but invariant only with respect to $\mathcal{G}_1$, not to the change of the origin. On the other hand, the proposed tests enjoy the properties (i), (iii) and (iv) mentioned above.

2.2. Liu and Singh rank sum test

An interesting test of Wilcoxon type, based on the ranks of depths of the data, was proposed by Liu and Singh [24]. Being of Wilcoxon type, this test is locally most powerful against some alternatives of the Lehmann type. Its asymptotic distributions under the hypothesis and under general alternative distributions F, G of depths was derived by Zuo and He [43].

Let $D(\mathbf{y}; H)$ denote a depth function of a distribution H evaluated at the point $\mathbf{y} \in \mathbb{R}^p$. Liu and Singh [24] considered a parameter, called a *quality index*, defined as

$$\begin{aligned} Q\big(F^{(p)}, G^{(p)}\big) &= \int R\big(\mathbf{y}; F^{(p)}\big)\, \mathrm{d}G^{(p)}(\mathbf{y}) \\ &= \mathbb{P}\big\{D\big(\mathbf{X}; F^{(p)}\big) \le D\big(\mathbf{Y}; F^{(p)}\big) | \mathbf{X} \sim F^{(p)}, \mathbf{Y} \sim G^{(p)}\big\}, \end{aligned}$$

where $R(\mathbf{y}; F^{(p)}) = \mathbb{P}_F(D(\mathbf{X}; F^{(p)}) \le D(\mathbf{y}; F^{(p)}), \mathbf{y} \in \mathbb{R}^p$, and showed that if $D(\mathbf{X}; F^{(p)})$ has a continuous distribution, then $Q(F^{(p)}, F^{(p)}) = \frac{1}{2}$. They then tested the hypothesis $Q(F^{(p)}, G^{(p)}) = \frac{1}{2}$ against the alternative $Q(F^{(p)}, G^{(p)}) \neq \frac{1}{2}$ using the Wilcoxon-type criterion based on the empirical distribution functions F_m and G_n of samples of sizes m and n respectively:

$$Q(F_m, G_n) = \int R(\mathbf{y}; F_m)\, \mathrm{d}G_n(\mathbf{y}) = \frac{1}{n} \sum_{j=1}^{n} R(\mathbf{Y}_j; F_m).$$

If the distribution of depths is symmetric under $F^{(p)} \equiv G^{(p)}$, then the test rejecting provided $|Q(F_m, G_n) - \frac{1}{2}| \ge C_{\alpha/2}$ is locally unbiased against $Q(F^{(p)}, G^{(p)}) \neq \frac{1}{2}$. Under a general distribution of depths, only the one-sided test with the critical region

$$Q(F_m, G_n) - \tfrac{1}{2} > C_\alpha$$

is unbiased against the one-sided alternative $Q(F^{(p)}, G^{(p)}) > \frac{1}{2}$; however, this alternative, one-sided in depths, has a difficult interpretation with respect to the distributions $F^{(p)}$ and $G^{(p)}$ of original observations $\mathbf{X}$ and $\mathbf{Y}$, respectively. Generally, the test is not finite-sample unbiased against $F \neq G$, not even locally. The unbiasedness can be guaranteed only in some cases, for instance, if the hypothetical distribution of depths is symmetric.

2.3. Rank tests based on distances of observations

We shall test the hypothesis of equality of distributions of two samples against alternatives that some distances are greater than others; because such alternatives are one-sided, they make the tests unbiased.

Choose a distance $L = L(\cdot, \cdot)$ in $\mathbb{R}^p$ taking nonnegative real values. Let

$$\mathcal{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = (\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

denote the pooled sample, where $N = m + n$, and consider the matrix of distances $\mathbb{L}_N = [\ell_{ik}]_{i,k=1}^N$, where $\ell_{ik} = L(\mathbf{Z}_i, \mathbf{Z}_k)$. We can construct simple rank tests based on $\mathbb{L}_N$ in three different ways:

(i) *Simple rank test, but not invariant with respect to $\mathcal{G}$ or $\mathcal{G}_1$.* Consider the vector

$$(\tilde{\ell}_1, \ldots, \tilde{\ell}_N), \qquad \tilde{\ell}_k = L(\mathbf{0}, \mathbf{Z}_k), \qquad k = 1, \ldots, N,$$

of distances of observations from the origin. The vector $(\tilde{\ell}_1, \ldots, \tilde{\ell}_m)$ is then a random sample from a population with a distribution function F (say), while $(\tilde{\ell}_{m+1}, \ldots, \tilde{\ell}_N)$ is a random sample from a population with a distribution function G. Assume that the distribution functions F and G are absolutely continuous. Under hypothesis $\mathbf{H}_0 : F^{(p)} \equiv G^{(p)}$, the distribution functions F and G coincide, that is, they satisfy the hypothesis $\widetilde{\mathbf{H}}_0 : F \equiv G$ which states that $\{\tilde{\ell}_k, k = 1, \ldots, N\}$ satisfy the hypothesis of randomness. If $\widetilde{\mathbf{H}}_0$ is not true, then $\mathbf{H}_0$ is not true either.

Denote by $(\tilde{R}_1, \dots, \tilde{R}_N)$ the respective ranks of $\{\tilde{\ell}_k, k = 1, \dots, N\}$. Under the hypothesis, the vector of ranks has the uniform distribution on the set of permutations of the numbers $1, \dots, N$. Because $\{\tilde{\ell}_k, k = 1, \dots, m\}$ and $\{\tilde{\ell}_k, k = m+1, \dots, N\}$ are random samples, under the hypothesis, as well as under the alternatives, every two-sample rank test will depend only on the ordered ranks $R^{(m+1)} < \dots < R^{(N)}$ of the second sample. However, although invariant with respect to increasing continuous functions of $(\tilde{\ell}_1, \dots, \tilde{\ell}_N)$, such a test would not be invariant with respect to the groups of transformations (2.1) or (2.3), even if $\tilde{\ell}_k = \|\mathbf{Z}_k\|$ is the Euclidean distance.

The linear rank test is based on the linear rank statistic

$$S_N = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_{ik}) \tag{2.6}$$

with the scores $a_N(1), \dots, a_N(N)$ generated by a nondecreasing score function $\varphi : (0, 1) \mapsto \mathbb{R}$ in either of the following two ways:

$$a_N(k) = \mathbb{E}(\varphi(U_{N:k})), \qquad k = 1, \dots, N, \tag{2.7}$$

or

$$a_N(k) = \varphi\left(\frac{U_{N:k}}{N+1}\right), \qquad k = 1, \dots, N, \tag{2.8}$$

with $U_{N:1} \le \dots \le U_{N:N}$ being the order statistics of the sample from the uniform $\mathbf{R}(0, 1)$ distribution. The test based on (2.6) is distribution-free, that is, the null distribution of S_N does not depend on the unknown $F \equiv G$ under $\widetilde{\mathbf{H}}_0$. Its asymptotic properties follow from [11] or [12].

(ii) *Conditional rank test, invariant with respect to* $\mathcal{G}_1$. Assuming that $m > p$, choose a suitable basis $(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_p}) = \mathbb{X}_p$ of $\{\mathbf{X}_i, 1 \le i \le m\}$; the choice of basis $\mathbb{X}_p$ can follow various aspects. Consider the set of $(m + n - p) \times p$ distances

$$\{\ell^*_{i_j,k} = L(\mathbf{X}_{i_j}, \mathbf{Z}_k), k = 1, \dots, N, k \ne i_1, \dots, i_p\}, \qquad j = 1, \dots, p.$$

Then, for a fixed $i_j, 1 \le j \le p$, and conditionally given $\mathbb{X}_p$, the vector $\{\ell^*_{i_j,k}, k = 1, \dots, m, k \ne i_1, \dots, i_p\}$ is a random sample from a population with a distribution function $F(z|\mathbb{X}_p) = F$ (say), while $\{\ell^*_{i_j,k}, k = m+1, \dots, N\}$ is a random sample from a population with a distribution function $G(z|\mathbb{X}_p) = G$. Assume that the distribution functions F and G are absolutely continuous. Let

$$\mathcal{R}_{i_j} = (R_{i_j,k}, k = 1, \dots, N, k \ne i_1, \dots, i_p)$$

denote the ranks of $\ell^*_{i_jk}, k = 1, \dots, N, k \ne i_j, \forall j = 1, \dots, p$. Every two-sample rank test will depend only on the ordered ranks $R^{(m+1)}_{i_j} < \dots < R^{(N)}_{i_j}$ of the second sample. In particular, if $L(\mathbf{X}_{i_j}, \mathbf{Z}_k) = \|\mathbf{X}_{i_j} - \mathbf{Z}_k\|, k = 1, \dots, N, k \ne i_1, \dots, i_p$, where $\|\cdot\|$ is the Euclidean distance, then the test based on their ranks will be invariant with respect to $\mathcal{G}_1$ in (2.4), but not with respect to $\mathcal{G}, \mathcal{G}_0$.

Similarly as in (i), the linear (conditional) rank test is based on the linear rank statistic

$$S^*_{i_j,N} = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_{i_j,k})$$

with the scores $a_N(1), \dots, a_N(N-p)$ generated by a nondecreasing score function φ as in either (2.7) or (2.8). The criteria $S^*_{i_j,N}$ are equally distributed for $j = 1, \dots, p$ under the hypothesis and under the alternatives, and are conditionally independent given $\mathbb{X}_p$. Using only a single $S^*_{i_j,N}$ would be a loss of information, so we look for a convenient combination of $S^*_{i_1,N}, \dots, S^*_{i_p,N}$. Every convenient homogeneous combination of $S^*_{i_1,N}, \dots, S^*_{i_p,N}$ leads to a rank test, conditional under given $\mathbb{X}_p$, which is distribution-free under the hypothesis. The problem would be to find its null distribution, and thus the critical values, under a finite N. The test based on a single $S^*_{i_j,N}$ is a standard rank test, for example, Wilcoxon, conditionally given $\mathbb{X}_p$, and is thus easy to perform. When we look for a similarly simple test based on a combination of $S^*_{i_jN}$, $1 \le j \le p$, it seems that the simplest possibility is a randomization of $S^*_{i_1,N}, \dots, S^*_{i_p,N}$, leading to the following criterion $\widetilde{S}^{(N)}$:

$$\mathbb{P}\big(\widetilde{S}^{(N)} = S^*_{i_j,N}\big) = \frac{1}{p}, \qquad j = 1, \dots, p, \tag{2.9}$$

where the randomization in (2.9) is independent of the set of observations $\mathcal{Z}$. The following identity is true for any C:

$$\mathbb{P}\big(\widetilde{S}^{(N)} > C\big) = \frac{1}{p} \sum_{j=1}^{p} \mathbb{P}(S^*_{i_j,N} > C)$$

and the test rejects $\widetilde{\mathbf{H}}_0$ for $\alpha \in (0,1)$ if $\widetilde{S}^{(N)} > C_\alpha$; eventually, it rejects with probability $\gamma \in (0,1)$ if $S^{(N)} = C_\alpha$, where

$$\mathbb{P}_{\widetilde{\mathbf{H}}_0}\big(\widetilde{S}^{(N)} > C_\alpha\big) + \gamma \mathbb{P}_{\widetilde{\mathbf{H}}_0}\big(\widetilde{S}^{(N)} = C_\alpha\big) = \alpha.$$

(iii) *Randomized rank test, invariant with respect to $\mathcal{G}_1$.* Similarly, for every fixed i and under fixed $\mathbf{X}_i$, $1 \le i \le m$, we can consider the distances $\{\ell^*_{ik} = L(\mathbf{X}_i, \mathbf{Z}_k), k = 1, \dots, N, k \ne i\}$. Then, conditionally given $\mathbf{X}_i$, the vector $\{\ell^*_{ik}, k = 1, \dots, m, k \ne i\}$ is a random sample from a population with a distribution function $F(z|\mathbf{X}_i) = F$ (say), while $\{\ell^*_{ik}, k = m+1, \dots, N\}$ is a random sample from a population with a distribution function $G(z|\mathbf{X}_i) = G$. Assuming that the distribution functions F and G are absolutely continuous, we work with

$$\mathcal{R}_i = (R_{i1}, \dots, R_{i,i-1}, R_{i,i+1}, \dots, R_{iN}),$$

the ranks of ℓ^*_{ik}, $k = 1, \dots, N$, $k \ne i$.

The linear (conditional) rank test is based on the linear rank statistic

$$S_{iN} = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_{ik}) \tag{2.10}$$

with the scores $a_N(1), \dots, a_N(N-1)$.

The criteria S_{iN} are equally distributed for $i = 1, \dots, m$ under the hypothesis and under the alternatives, although not independent. We look for a convenient combination of $S_{1N}, \dots, S_{mN}$. Again, a randomization of $S_{1N}, \dots, S_{mN}$ keeps the simple structure of the test and is thus easy to perform. It leads to the following criterion, $S^{(N)}$:

$$\mathbb{P}\big(S^{(N)} = S_{iN}\big) = \frac{1}{m}, \qquad i = 1, \dots, m, \tag{2.11}$$

where the randomization in (2.11) is independent of the set of observations $\mathcal{Z}$. Again, for any C,

$$\mathbb{P}\big(S^{(N)} > C\big) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{P}(S_{iN} > C),$$

and the test rejects $\widetilde{\mathbf{H}}_0$ for $\alpha \in (0,1)$ if $S^{(N)} > C_\alpha$; eventually, it rejects with probability $\gamma \in (0,1)$ if $S^{(N)} = C_\alpha$. Again, with the Euclidean distance, the test will be invariant with respect to $\mathcal{G}_1$ in (2.4), but not with respect to $\mathcal{G}, \mathcal{G}_0$.

Remark 2.1. The Mahalanobis distances

$$\mathbf{Z}_k^\top (\mathbf{V}_N^0)^{-1} \mathbf{Z}_k, \qquad k = 1, \dots, N, \tag{2.12}$$

$$(\mathbf{X}_i - \mathbf{Z}_k)^\top \mathbf{V}_N^{-1} (\mathbf{X}_i - \mathbf{Z}_k) \quad \text{or} \quad (\mathbf{X}_i - \mathbf{Z}_k)^\top (\mathbf{V}_N^0)^{-1} (\mathbf{X}_i - \mathbf{Z}_k), \qquad k \neq i, \tag{2.13}$$

are not independent, but under $\mathbf{H}_0$, they have exchangeable distributions; hence, under $\mathbf{H}_0$, the distribution of their ranks is independent of the distribution of observations (is distribution-free). Moreover, (2.13) are invariant with respect to $\mathcal{G}$ and $\mathcal{G}_0$, while (2.12) are invariant only with respect to $\mathcal{G}_0$. The invariant tests based on the ranks of (2.12) or (2.13) will be the subject of a further study. Their structure is more complex than that of tests based on simple distances.

3. Structure of the rank tests

Let $\mathbf{X} = (X_1, \dots, X_m)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ be two independent samples from distributions F and G, respectively. Consider the rank test with the criterion $S_N = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_i)$, where $R_1, \dots, R_N$ are the ranks of the pooled sample $\mathbf{Z} = (X_1, \dots, X_m, Y_1, \dots, Y_n)$. The values X_i, Y_j are, for example, the distances of multivariate observations, either from a fixed point or the interpoint distances considered conditionally given the original component. We want to test

the hypothesis $\mathbf{H}_0 : F \equiv G$ against a general alternative with the $(m+n)$-dimensional distribution function of the form

$$\mathbf{K}: \prod_{k=1}^{m} G_{\Delta}^{(1)}(z_k) \prod_{k=m+1}^{N} G_{\Delta}^{(2)}(z_k). \tag{3.1}$$

Lehmann [20] showed that the Wilcoxon two-sample test with the score-generating function $\varphi(u) = 2u - 1, 0 \le u \le 1$, is the locally most powerful rank test of $\mathbf{H}_0$ against the class of alternatives with

$$G_{\Delta}^{(1)}(z) = F(z) \quad \text{and} \quad G_{\Delta}^{(2)}(z) = G_{\Delta}(z), \tag{3.2}$$

$$G_{\Delta}(z) = \begin{cases} (1-\Delta)F(z) + \Delta F^2(z), & z \ge 0, \\ 0, & z < 0 \end{cases} \tag{3.3}$$

with $0 < \Delta < 1$. The Y's are then stochastically larger than the X's and $F(z) - G_{\Delta}(z) \equiv \Delta \cdot F(z)(1 - F(z))$; hence, the Kolmogorov distance of F and G_{Δ} is

$$d_K(F, G_{\Delta}) = \Delta \cdot \sup_{z \ge 0}\big[F(z)\big(1 - F(z)\big)\big] = \frac{\Delta}{4}$$

and the point of maximum is $z = F^{-1}(\frac{1}{2})$.

Gibbons [9] proved that the *Psi test* with the scores

$$a_N(i) = \sum_{j=0}^{i-1} \frac{1}{N-j} - \sum_{j=0}^{N-i} \frac{1}{N-j}, \qquad i = 1, \ldots, N,$$

is the locally most powerful rank test of $\mathbf{H}_0$ against the alternative (3.1) with

$$\begin{aligned} G_{\Delta}^{(1)}(z) &= 1 - \big(1 - F(z)\big)^{1+\Delta}, \\ G_{\Delta}^{(2)}(z) &= (F(z))^{1+\Delta}, \qquad \Delta > 0,\ z \ge 0. \end{aligned} \tag{3.4}$$

Obviously,

$$\begin{aligned} G_{\Delta}^{(1)}(z) &\ge F(z) \ge G_{\Delta}^{(2)}(z) \\ &= (F(z))^{1+\Delta} \qquad \forall z \ge 0 \text{ and } \Delta \ge 0, \end{aligned}$$

hence $G_{\Delta}^{(1)}$ is stochastically smaller than $G_{\Delta}^{(2)}$ for $\Delta > 0$. The Kolmogorov distance of $G_{\Delta}^{(1)}$ and $G_{\Delta}^{(2)}$ is

$$\begin{aligned} d_K\big(G_{\Delta}^{(1)}, G_{\Delta}^{(2)}\big) &= \sup_{z \ge 0}\big[1 - (F(z))^{1+\Delta} - \big(1 - F(z)\big)^{1+\Delta}\big] \\ &= 1 - 2^{-\Delta} \end{aligned}$$

and the maximum is attained at the point $z = F^{-1}(\frac{1}{2})$. The score generating function of the Psi test is $\varphi(u) = \ln u - \ln(1-u), 0 < u < 1$.

Similarly, Savage [36] proved that the *Savage test* with the critical region

$$\sum_{i=1}^{n} \sum_{j=R_{m+i}}^{N} \frac{1}{j} \le C_\alpha$$

is the locally most powerful rank test of $\mathbf{H}_0$ against the class of alternatives (3.1) with

$$G_\Delta^{(1)}(z) = F(z), \qquad G_\Delta^{(2)}(z) = F^{1+\Delta}(z), \qquad z \ge 0,\ \Delta > 0. \tag{3.5}$$

Again, the Y's are stochastically larger than the X's, and the Kolmogorov distance of $G_\Delta^{(1)}$ and $G_\Delta^{(2)}$ is equal to

$$\begin{aligned} d_K\big(G_\Delta^{(1)}, G_\Delta^{(2)}\big) &= \sup_{z\ge 0}\big[F(z)\big(1 - F^\Delta(z)\big)\big] \\ &= \Delta(1+\Delta)^{-1-1/\Delta}, \end{aligned}$$

whose maximum is attained at $z = F^{-1}((1+\Delta)^{-1/\Delta})$. The score generating function of this test is $\varphi(u) = 1 + \ln u, 0 < u < 1$.

Assume that F is increasing and let $U_k = F(z_k), k = 1, \dots, N$. Under the alternative (3.2), the ranks $R_1, \dots, R_N$ are also the ranks of the variables $U_1, \dots, U_m, V_{m+1}, \dots, V_N$, where $V_k = (1-\Delta)U_k + \Delta U_k^2, k = m+1, \dots, N$. An analogous consideration applies to the alternatives (3.4) and (3.5). Hence, the distribution of the ranks $R_1, \dots, R_N$ is independent of F (is distribution-free) under the hypothesis as well as under the alternatives, and thus the power functions of all rank tests against the alternatives (3.2), (3.4) and (3.5) are distribution-free. The Lehmann alternatives can be well interpreted, are flexible and can describe various experimental situations well. Besides the linear rank tests, we can also consider the two-sample Kolmogorov–Smirnov test based on the empirical distribution functions of the interpoint distances, for the purposes of comparison. The randomized Kolmogorov–Smirnov test, following a similar structure as the tests in Section 2, is also distribution-free. Instead of interpoint distances, we can consider the rank tests based on the depths using similar ideas.

We shall concentrate on the two-sample Wilcoxon, Psi and Savage rank tests because they are easy to perform, are locally most powerful and are locally unbiased against some alternatives of Lehmann type. The ranks are distribution-free not only under the hypothesis, but also under the Lehmann alternatives, hence the powers of the rank tests are independent of the distribution of the data. This is an advantage because we do not need to calculate the distribution of the distances. Several authors (e.g., [8,16,25,35,36]) considered various distances of two sets of multivariate observations from some specified point, constructed the critical regions and verified their consistencies against distant alternatives. However, the questions of the finite-sample behavior of these tests, their unbiasedness and against which alternatives, and their efficiency against local alternatives, remains open. If the test is not unbiased against some alternative of interest, then its power can be less than the significance level, say less than $\alpha = 0.05$, hence such a test is not suitable for verifying the hypothesis against this specific alternative.

The sequences of alternatives (3.1) corresponding to (3.3), (3.4) and (3.5) are contiguous with respect to the sequence $\{\prod_{i=1}^{N} F(z_i)\}$ provided that $\Delta_N = N^{-1/2}\Delta_0$ with Δ_0 fixed, $0 < \Delta_0 < \infty$, as shown in the Appendix. Hence, we are able to evaluate the local asymptotic powers of the tests; this is done in the next section, along with the numerical illustration and comparison of the tests.

3.1. Local asymptotic powers of the tests

We shall assume throughout that

$$\lim_{N\to\infty} \frac{m_N}{N} = \lambda \in (0, 1).$$

Let $U_i = F(Z_i)$, $i = 1, \dots, N$. The alternative (3.1) in the special cases (3.2), (3.4) and (3.5) can then be rewritten as follows:

$$\begin{aligned}
&\widetilde{G}_\Delta^{(1)}(u) = u, \qquad \widetilde{G}_\Delta^{(2)}(u) = (1-\Delta)u + \Delta u^2, \qquad 0 \le u \le 1,\\
&\widetilde{G}_\Delta^{(1)}(u) = 1 - (1-u)^{1+\Delta}, \qquad \widetilde{G}_\Delta^{(2)}(u) = \widetilde{G}(u, \Delta) = u^{1+\Delta},\\
&\widetilde{G}_\Delta^{(1)}(u) = u, \qquad \widetilde{G}_\Delta^{(2)}(u) = \widetilde{G}(u, \Delta) = u^{1+\Delta},\\
&\qquad \Delta > 0, \qquad 0 \le u \le 1.
\end{aligned} \tag{3.6}$$

Because Δ is the parameter of interest and the alternatives (3.6) are contiguous with respect to the sequence of hypotheses $\{\prod_{i=1}^{N} u_i I[0 \le u_i \le 1]\}$ under $\Delta_N = N^{-1/2}\Delta_0$ (see Appendix for the proof), we can study the powers of the rank tests under alternatives (3.6) without loss of generality.

Consider the centered test criterion

$$S_N^* = N^{-1/2}\left[-\frac{n}{N}\sum_{i=1}^{m} a_N(R_{Ni}) + \frac{m}{N}\sum_{i=m+1}^{N} a_N(R_{Ni})\right]. \tag{3.7}$$

If the scores are generated by a nondecreasing score function $\varphi : (0, 1) \mapsto \mathbb{R}$ which is square-integrable on $(0, 1)$, then the asymptotic distribution of (3.7) under contiguous alternatives follows from the LeCam theorems (see [11] or [12]). Namely, S_N^* will be asymptotically normally distributed $\mathcal{N}(\mu, \sigma^2)$ with

$$\mu = \lambda(1-\lambda)\int_0^1 \varphi(u)\varphi^*(u)\,\mathrm{d}u, \qquad \varphi^*(u) = \left.\frac{\partial \ln \tilde{g}(u, \Delta)}{\partial \Delta}\right|_{\Delta=0},$$

$$\tilde{g}(u, \Delta) = \frac{\partial \tilde{G}(u, \Delta)}{\partial u} \qquad \text{being the density of } \widetilde{G}(u, \Delta),$$

$$\sigma^2 = \lambda(1-\lambda)\int_0^1 \varphi^2(u)\,\mathrm{d}u.$$

The test rejects $\mathbf{H}$ on the significance level α provided $S_N^* \geq \sigma\Phi^{-1}(1-\alpha)$, where Φ is the standard normal distribution function. Hence, the asymptotic power of the test under the alternative $\mathbf{K}_N : \Delta_N = N^{-1/2}\Delta_0$ equals

$$\begin{aligned}
&\lim_{m,n\to\infty} \mathbb{P}_{\mathbf{K}_N}\left(\frac{S_N^* - \mu}{\sigma} \geq \Phi^{-1}(1-\alpha) - \frac{\mu}{\sigma}\right) \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\mu}{\sigma}\right) \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{\lambda(1-\lambda)}}{A}\Delta_0 \int_0^1 \varphi(u)\varphi^*(u)\,\mathrm{d}u\right),
\end{aligned}$$

where $A^2 = \int_0^1 \varphi^2(u)\,\mathrm{d}u$. The relative asymptotic efficiency of a test S_{N1}^* with respect to a different test S_{N2}^* is given as the ratio

$$\left(\frac{\mu^{(1)}}{\sigma_1} \Big/ \frac{\mu^{(2)}}{\sigma_2}\right)^2, \tag{3.8}$$

where $\mu^{(1)}$ and σ_1^2 are, respectively, the asymptotic mean and variance of the statistic S_{N1}^* and $\mu^{(2)}$ and σ_2^2 are those of S_{N2}^*.

Table 1 summarizes the relative asymptotic efficiencies of the Wilcoxon, Psi and Savage tests with respect to the locally most powerful rank test for specified Lehmann alternatives. These values are computed with the aid of (3.8). For the purposes of illustration, we also add the van der Waerden and median tests, and their relative asymptotic efficiencies with respect to the locally most powerful rank tests.

For the next illustration, consider the Lehmann alternative (3.2) and compare the locally most powerful Wilcoxon test (the score function $\varphi(u) = 2u - 1, 0 \leq u \leq 1$) with the Kolmogorov–Smirnov test. The asymptotic power of the Wilcoxon test against $\mathbf{K}_N$ equals

$$\begin{aligned}
&\lim_{\min(m,n)\to\infty} \mathbb{P}\left(\left[\frac{1}{3}\lambda(1-\lambda)\right]^{-1/2} S_N \geq \Phi^{-1}(1-\alpha)\Big|\mathbf{K}_N\right) \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \Delta_0\sqrt{\frac{\lambda(1-\lambda)}{3}}\right).
\end{aligned} \tag{3.9}$$

Table 1. Relative asymptotic efficiencies under various alternatives

Alternative	Test				
	Wilcoxon	Psi	Savage	van der Waerden	Median
(3.2)	1.000	0.912	0.750	0.955	0.750
(3.4)	0.912	1.000	0.882	0.992	0.584
(3.5)	0.750	0.822	1.000	0.816	0.480

For small values Δ_0, it can be further approximated in the following way:

$$\begin{aligned}&\mathbb{P}\left(\left[\frac{1}{3}\lambda(1-\lambda)\right]^{-1/2} S_N \geq \Phi^{-1}(1-\alpha)\big|\mathbf{K}_N\right)\\&\approx \alpha + \Delta_0 \cdot \Phi^{\top}\left(\Phi^{-1}(1-\alpha)\right)\sqrt{\frac{\lambda(1-\lambda)}{3}}.\end{aligned} \tag{3.10}$$

Let us now consider the Kolmogorov–Smirnov test against alternative (3.2). Let $\hat{F}_m$ and $\hat{G}_n$ be the respective empirical distribution functions of samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. Then, by Hájek *et al.* [12], Theorem VI.3.2, we have

$$\begin{aligned}&\lim_{m,n\to\infty} \mathbb{P}\left(\sqrt{\frac{nm}{n+m}} \sup_{x\in R}\left(\hat{G}_n(x) - \hat{F}_m(x)\right) \geq \sqrt{-\frac{1}{2}\log\alpha}\,|\mathbf{K}_N\right)\\&= \mathbb{P}\left(\sup_{0\leq u\leq 1}\left(\mathcal{B}(u) + \Delta_0\sqrt{\lambda(1-\lambda)}u(1-u)\right) \geq \sqrt{-\frac{1}{2}\log\alpha}\right),\end{aligned}$$

where $\mathcal{B}(u)$ is a Brownian bridge. The last probability cannot easily be calculated analytically. Hence, we resort to a linear approximation around the point $\Delta_0 = 0$ and get

$$\begin{aligned}&\mathbb{P}\left(\sup_{0\leq u\leq 1}\left(\mathcal{B}(u) + \Delta_0\sqrt{\lambda(1-\lambda)}u(1-u)\right) \geq \sqrt{-\frac{1}{2}\log\alpha}\right)\\&\approx \alpha + 2\Delta_0\sqrt{\lambda(1-\lambda)}\alpha\sqrt{-\frac{1}{2}\log\alpha}\int_0^1 (2u-1)\psi(\alpha,u)\,\mathrm{d}u,\end{aligned} \tag{3.11}$$

where

$$\psi(\alpha,u) = 2\Phi\left(\frac{(2u-1)\sqrt{-(1/2)\log\alpha}}{\sqrt{u(1-u)}}\right) - 1.$$

Table 2 gives the asymptotic powers (for $\alpha = 0.05$) of the Wilcoxon test (As.W) and the Kolmogorov–Smirnov test (As.KS) computed from (3.9) and (3.11); these powers are compared with empirical powers (Obs.W, Obs.KS) obtained by simulations of 30, 100, 500 and 1000 observations in both samples. The simulations were carried out in the R programming language using 500 000 replications under the alternative (3.3), where F denotes the distribution function of the uniform $\mathbf{R}(0,1)$ distribution. We recall that the powers of rank tests under Lehmann alternatives are also distribution-free for finite samples.

The asymptotic approximation (3.9) of the power of the Wilcoxon test is already very good for $m = n = 100$. Unfortunately, the linear approximation (3.11) of the power of the Kolmogorov–Smirnov test only works in a local neighborhood of the null hypothesis as the power function increases exponentially. Even for small values of Δ_0, the approximation (3.11) of the power of the Kolmogorov–Smirnov test is very good only for large sample sizes.

Table 2. Comparison of the empirical powers for various sample sizes and of the local asymptotic powers of the Wilcoxon and Kolmogorov–Smirnov tests against the alternative (3.3)

Δ_0	Obs.W, $m = n =$				As.W	Obs.KS, $m = n =$				As.KS
	30	100	500	1000		30	100	500	1000	
0.0	0.050	0.050	0.050	0.050	0.050	0.036	0.039	0.048	0.048	0.050
0.1	0.052	0.053	0.053	0.053	0.053	0.038	0.040	0.050	0.052	0.053
0.2	0.056	0.056	0.056	0.056	0.056	0.040	0.043	0.053	0.054	0.055
0.3	0.059	0.059	0.059	0.059	0.060	0.042	0.046	0.056	0.057	0.058
0.4	0.063	0.064	0.063	0.063	0.063	0.044	0.050	0.059	0.060	0.061
0.5	0.067	0.066	0.067	0.066	0.067	0.047	0.052	0.063	0.063	0.063
0.6	0.069	0.069	0.071	0.071	0.071	0.049	0.055	0.066	0.067	0.066
0.7	0.076	0.074	0.074	0.075	0.075	0.053	0.057	0.069	0.070	0.069
0.8	0.077	0.079	0.079	0.080	0.079	0.054	0.061	0.073	0.074	0.072
0.9	0.081	0.083	0.083	0.082	0.083	0.057	0.063	0.076	0.077	0.074
1.0	0.085	0.088	0.087	0.088	0.088	0.060	0.067	0.080	0.081	0.077
2.0	0.141	0.142	0.141	0.143	0.143	0.100	0.107	0.126	0.131	0.104
3.0	0.214	0.217	0.215	0.218	0.218	0.155	0.165	0.185	0.193	0.131

Table 3 compares the slopes in linear approximations of asymptotic powers of the Kolmogorov–Smirnov and Wilcoxon tests, given in (3.10) and (3.11), under various sizes of the tests. The first column gives the size of the test, the second column the slope for the Kolmogorov–Smirnov test (K–S), the third column gives the slope for the Wilcoxon test and the last column gives the ratio of the two slopes.

4. Numerical comparison of Hotelling- and Wilcoxon-type tests

The empirical powers of the Hotelling T^2 and Wilcoxon two-sample tests are compared under bivariate normal and Cauchy distributions with various parameters; the Wilcoxon test of type (2.10), (2.11) is based on the ranks of the Euclidean interpoint distances. The Hotelling test distinguishes well two normal samples contrasting in locations, even if they also differ in scales.

Table 3. Slopes of the Kolmogorov–Smirnov and Wilcoxon tests at various levels of significance

α	K–S	Wilcoxon	Wilcoxon/K–S
0.001	0.001	0.002	2.070
0.010	0.009	0.015	1.680
0.025	0.022	0.034	1.500
0.050	0.044	0.059	1.350
0.100	0.086	0.101	1.180

Table 4. Powers of two-sample Hotelling T^2 test (H) and of two-sample Wilcoxon test (W) based on distances for various $m = n$ and $\alpha = 0.05$. The first sample always has $\mathcal{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ distribution with $\boldsymbol{\mu}_1 = (0, 0)^\top$ and $\boldsymbol{\Sigma}_1 = \mathrm{Diag}\{1, 1\}$. The second sample has $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ with various $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_2$ specified in the first column

Second sample	Test	$m = n = 10$	$m = n = 100$	$m = n = 1000$
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0471	0.0481	0.0493
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1, 1\}$	W	0.0457	0.0505	0.0487
$\boldsymbol{\mu}_2 = (0.2, 0.2)^T$	H	0.0771	0.4115	1.0000
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1, 1\}$	W	0.0520	0.1715	0.6458
$\boldsymbol{\mu}_2 = (0.5, 0.5)^T$	H	0.2318	0.9962	1.0000
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1, 1\}$	W	0.1085	0.5701	0.8617
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0659	0.0561	0.0452
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{0.1, 0.1\}$	W	0.7994	0.9998	1.0000
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0653	0.0456	0.0530
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{0.2, 0.2\}$	W	0.4851	0.9932	1.0000
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0521	0.0521	0.0463
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{0.5, 0.5\}$	W	0.1182	0.7034	0.9968
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0531	0.0530	0.0514
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1.5, 1.5\}$	W	0.0656	0.2881	0.8525
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0552	0.0518	0.0508
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{2, 2\}$	W	0.0999	0.5395	0.9670
$\boldsymbol{\mu}_2 = (0, 0)^T$	H	0.0572	0.0546	0.0521
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1.0, 0.2\}$	W	0.1029	0.6568	0.9936
$\boldsymbol{\mu}_2 = (0.1, 0.1)^T$	H	0.0553	0.1266	0.7897
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1.1, 1.1\}$	W	0.0491	0.0932	0.4232
$\boldsymbol{\mu}_2 = (0.1, 0.1)^T$	H	0.0601	0.1167	0.7183
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1.5, 1.5\}$	W	0.0667	0.3182	0.7690
$\boldsymbol{\mu}_2 = (0.2, 0.2)^T$	H	0.0742	0.3656	1.0000
$\sigma_1^2 = 1, \sigma_2^2 = 1.5$	W	0.0548	0.2246	0.6907
$\boldsymbol{\mu}_2 = (0.2, 0.2)^T$	H	0.0710	0.3402	0.9994
$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\{1.5, 1.5\}$	W	0.0668	0.3551	0.7597

However, in some situations, the Wilcoxon test even competes well with the Hotelling test, namely, when either the samples differ only moderately in locations or when they differ considerably in scales. This is illustrated by Table 4, which provides empirical powers of Hotelling and Wilcoxon tests for a comparison of two bivariate normal samples. The sample sizes are $m = n = 10, 100, 1000$ and the simulations are based on 10 000 replications. The first sample always has distribution $\mathcal{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{\mu}_1 = (0, 0)^\top$ and $\boldsymbol{\Sigma}_1 = \mathrm{Diag}\{1, 1\}$, while the second sample has $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ with various parameters.

We also refer to the simulation study of [43] which compared the empirical powers of the Liu–Singh rank-sum test (Q) based on the depths, the Hotelling and the Hetmansperger *et al.* [17] tests for two bivariate normal samples. Under normality, the Q-test mostly dominates the

Table 5. Powers of two-sample Hotelling T^2 test (H) and two-sample Wilcoxon test (W) based on distances for various $m = n$ and $\alpha = 0.05$. The first sample $\mathbf{X}$ always has the two-dimensional Cauchy distribution. The second sample $\mathbf{Y}$ is obtained as $\mathbf{Y} = \boldsymbol{\mu} + \sigma \mathbf{Y}^*$, where $\mathbf{Y}^*$ is generated as a two-dimensional Cauchy distribution independent of $\mathbf{X}$. Values of $m, n, \boldsymbol{\mu}$ and σ are specified in the first column

Second sample	Test	$m = n = 10$	$m = n = 25$	$m = n = 100$	$m = n = 1000$
$\boldsymbol{\mu} = (0, 0)^T$	H	0.0191	0.0156	0.0171	0.0217
$\sigma = 1$	W	0.0450	0.0478	0.0510	0.0442
$\boldsymbol{\mu} = (0.2, 0.2)^T$	H	0.0227	0.0232	0.0227	0.0174
$\sigma = 1$	W	0.0468	0.0536	0.0874	0.3925
$\boldsymbol{\mu} = (0.5, 0.5)^T$	H	0.0408	0.0404	0.0414	0.0361
$\sigma = 1$	W	0.0664	0.1115	0.2937	0.7470
$\boldsymbol{\mu} = (1, 1)^T$	H	0.1038	0.1193	0.1260	0.1226
$\sigma = 1$	W	0.1219	0.2710	0.6235	0.8893
$\boldsymbol{\mu} = (5, 5)^T$	H	0.7387	0.7535	0.7683	0.7772
$\sigma = 1$	W	0.7574	0.9441	0.9782	0.9944
$\boldsymbol{\mu} = (0, 0)^T$	H	0.0200	0.0171	0.0193	0.0103
$\sigma = 1.5$	W	0.0664	0.1207	0.3419	0.8428
$\boldsymbol{\mu} = (0, 0)^T$	H	0.0207	0.0168	0.0182	0.0172
$\sigma = 2$	W	0.1082	0.2439	0.6123	0.9135
$\boldsymbol{\mu} = (0.2, 0.2)^T$	H	0.0189	0.0201	0.0196	0.0240
$\sigma = 1.5$	W	0.0710	0.1297	0.3495	0.8249
$\boldsymbol{\mu} = (1, 1)^T$	H	0.0741	0.0814	0.0865	0.0943
$\sigma = 1.5$	W	0.1088	0.2188	0.4925	0.8462
$\boldsymbol{\mu} = (2, 2)^T$	H	0.2356	0.2546	0.2690	0.2716
$\sigma = 1.5$	W	0.2092	0.4139	0.7395	0.9259
$\boldsymbol{\mu} = (0.2, 0.2)^T$	H	0.0248	0.0186	0.0217	0.0158
$\sigma = 2$	W	0.1134	0.2401	0.5990	0.9151
$\boldsymbol{\mu} = (1, 1)^T$	H	0.0575	0.0616	0.0623	0.0676
$\sigma = 2$	W	0.1330	0.2797	0.5619	0.8272
$\boldsymbol{\mu} = (2, 2)^T$	H	0.1796	0.1936	0.2045	0.2164
$\sigma = 2$	W	0.1771	0.3513	0.6531	0.8981

other two tests, as well as the Wilcoxon test based on interpoint distances. However, the (local) unbiasedness of the Q-test against two-sample alternatives is doubtful under asymmetric distributions of the depths, while a one-sided alternative in depths has a difficult interpretation in the original data.

Table 5 presents the empirical powers of the tests comparing two samples from the bivariate Cauchy distributions. The first sample $\mathbf{X}$ has a two-dimensional Cauchy distribution with independent components. The second sample $\mathbf{Y}$ is obtained as a random sample $\mathbf{Y}^*$ from the two-dimensional Cauchy distribution with independent components, independent of $\mathbf{X}$, transformed to $\mathbf{Y} = \boldsymbol{\mu} + \sigma \mathbf{Y}^*$ for certain shifts $\boldsymbol{\mu}$ and scales σ. The results are based on 10 000 replications.

The Wilcoxon test is far more powerful than the Hotelling test, already under a small shift. The Hotelling test fails completely if $\boldsymbol{\mu} = 0$ but $\sigma \neq 1$, while the Wilcoxon test still distinguishes well the samples. The Wilcoxon test also dominates the Hotelling test in other situations.

The rank tests based on interpoint distances are distribution-free, both under the hypothesis and under the Lehmann alternatives, while the exact distribution of the distances can remain unknown when performing the tests. The tests are locally unbiased against one-sample alternatives. If the interpoint distances are replaced with other scalar characteristics which are symmetrically distributed under the hypothesis, then the tests are also locally unbiased against the two-sample alternatives. The Lehmann alternatives reflect the practical situations well.

Appendix: Contiguity of Lehmann's alternatives

Let $\{P_{N1}, \dots, P_{NN}\}$ and $\{Q_{N1}, \dots, Q_{NN}\}$ be two triangular arrays of probability measures defined on the measurable space $(\mathcal{X}, \mathcal{A})$, and let $P_N^{(N)} = \prod_{i=1}^N P_{Ni}$ and $Q_N^{(N)} = \prod_{i=1}^N Q_{Ni}$ denote the respective product measures, $N = 1, 2, \dots$. Further, denote by p_{Ni} and q_{Ni} the respective densities of P_{Ni} and Q_{Ni} with respect to a σ-finite measure μ_i, which can also be $\mu_i = P_{Ni} + Q_{Ni}, i = 1, \dots, N$.

Oosterhoff and van Zwet [32] proved that $\{Q_N^{(N)}\}$ is contiguous with respect to $\{P_N^{(N)}\}$ if and only if

$$\limsup_{N\to\infty} \sum_{k=1}^N H^2(P_{Nk}, Q_{Nk}) < \infty \tag{A.1}$$

and

$$\lim_{N\to\infty} \sum_{k=1}^N Q_{Nk}\left\{\frac{q_{Nk}(X_{Nk})}{p_{Nk}(X_{Nk})} \geq c_N\right\} = 0 \qquad \forall c_N \to \infty, \tag{A.2}$$

where

$$H(P, Q) = \left[\int \left(\sqrt{p} - \sqrt{q}\right)^2 \mathrm{d}\mu\right]^{1/2} = \left[2\int \left(1 - \sqrt{pq}\right) \mathrm{d}\mu\right]^{1/2}$$

is the Hellinger distance of P, Q.

Put $\Delta_N = N^{-1/2}\Delta_0$ with $\Delta_0 > 0$ fixed. Applying (A.1) and (A.2), we can verify the contiguity of the sequence $\{\prod_{k=1}^m G_{\Delta_N}^{(1)}(z_k) \prod_{k=m+1}^N G_{\Delta_N}^{(2)}(z_k)\}$ with respect to $\{\prod_{k=1}^N F(z_k)\}$ for the alternatives (3.3), (3.4) and (3.5).

Lemma A.1. (i) *Let*

$$\left\{\prod_{k=1}^N F(z_k)\right\}_{N=1}^{\infty}, \qquad z_1, \dots, z_N \geq 0,\ N = 1, 2, \dots, \tag{A.3}$$

and

$$\left\{\prod_{k=1}^{m} G_{\Delta_N}^{(1)}(z_k) \prod_{k=m+1}^{N} G_{\Delta_N}^{(2)}(z_k)\right\}_{N=1}^{\infty}, \qquad z_1,\dots,z_N \geq 0,\ N=1,2,\dots, \tag{A.4}$$

be two sequences of probability distributions satisfying

$$\Delta_N = N^{-1/2}\Delta_0 > 0, \qquad \lim_{N\to\infty} \min\{m,n\} = \infty,$$
$$\lim_{N\to\infty} \frac{m}{N} = \lim_{N\to\infty} \frac{m_N}{N} = \lambda \in (0,1),$$

where $G_\Delta^{(1)}, G_\Delta^{(2)}$ *are given by either* (3.2), (3.4) *or* (3.5). *The sequence* (A.4) *is then contiguous with respect to the sequence* (A.3).

Proof. (i) Let us first consider the Lehmann alternatives (3.2). Then,

$$\begin{aligned}
&\sum_{k=m+1}^{N} H^2(F(z_k), G_{\Delta_N}(z_k)) \\
&\quad = n \cdot \int_0^\infty f(z)\Big[\sqrt{1+\Delta_N\big(2F(z)-1\big)}-1\Big]^2 \,\mathrm{d}z \\
&\quad = n \int_0^\infty f(z) \frac{[1+\Delta_N(2F(z)-1)-1]^2}{[\sqrt{1+\Delta_N(2F(z)-1)}+1]^2}\,\mathrm{d}z \\
&\quad \leq n\Delta_N^2 \int_0^\infty f(z)\big(2F(z)-1\big)^2\,\mathrm{d}z \\
&\quad = 4n\Delta_N^2 \int_0^1 \left(u-\frac{1}{2}\right)^2 = \frac{1}{3}n\Delta_N^2 = \frac{1}{3}\lambda_N\Delta_0 < \infty
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{N\to\infty} \sum_{k=1}^{N} Q_{Nk}\left\{\frac{q_{Nk}(X_{Nk})}{p_{Nk}(X_{Nk})} \geq c_N\right\} \\
&\quad = 0 + \lim_{N\to\infty} \sum_{k=m+1}^{N} \int_0^\infty I\big[1+\Delta_N\big(2F(z_k)-1\big) \geq c_N\big] \\
&\qquad\qquad \times \big[1+\Delta_N\big(2F(z_k)-1\big)\big] f(z_k)\,\mathrm{d}z_k = 0
\end{aligned}$$

because $c_N > 1 + N^{-1/2}\Delta_0$ for $n > N_0$ whenever $c_N \to \infty$. The contiguity is thus verified.

(ii) Let

$$G_{\Delta_N}(z_i) = \begin{cases} 1-\big(1-F(z_i)\big)^{1+\Delta_N}, & i \leq m, \\ (F(z_i))^{1+\Delta_N}, & i \geq m+1. \end{cases}$$

Then,

$$
\begin{aligned}
&\sum_{i=1}^{N} H^2(F(z_i), G_{\Delta_N}(z_i)) \\
&\quad = m \cdot \int_0^\infty f(z)\Big[\sqrt{(1+\Delta_N)\big(1-F(z)\big)^{\Delta_N}} - 1\Big]^2 \,\mathrm{d}z \\
&\qquad + n \cdot \int_0^\infty f(z)\Big[\sqrt{(1+\Delta_N)(F(z))^{\Delta_N}} - 1\Big]^2 \,\mathrm{d}z \\
&\quad \le m \cdot \int_0^\infty f(z)\Big[(1+\Delta_N)\big(1-F(z)\big)^{\Delta_N} - 1\Big]^2 \,\mathrm{d}z \\
&\qquad + n \cdot \int_0^\infty f(z)[(1+\Delta_N)(F(z))^{\Delta_N} - 1]^2 \,\mathrm{d}z \\
&\quad = m \cdot \int_0^1 [(1+\Delta_N)(1-u)^{\Delta_N} - 1]^2 \,\mathrm{d}z + n \cdot \int_0^1 [(1+\Delta_N)u^{\Delta_N} - 1]^2 \,\mathrm{d}z \\
&\quad = N \cdot \int_0^1 [(1+\Delta_N)u^{\Delta_N} - 1]^2 \,\mathrm{d}z \le \Delta_N < \infty
\end{aligned}
$$

and hence (A.1) is proved for the alternative (3.4). Concerning (A.2), we have

$$
\begin{aligned}
&\lim_{N\to\infty} \sum_{i=1}^{N} Q_{Ni}\left\{\frac{q_{Ni}(X_{Ni})}{p_{Ni}(X_{Ni})} \ge c_N\right\} \\
&\quad = \lim_{N\to\infty}\Bigg\{m \cdot \int_0^\infty I\Big[(1+\Delta_N)\big(1-F(z)\big)^{\Delta_N} \ge c_N\Big](1+\Delta_N)\big(1-F(z)\big)^{\Delta_N} f(z)\,\mathrm{d}z \\
&\qquad\qquad + n \cdot \int_0^\infty I[(1+\Delta_N)(F(z))^{\Delta_N} \ge c_N](1+\Delta_N)(F(z))^{\Delta_N} f(z)\,\mathrm{d}z\Bigg\} \\
&\quad = \lim_{N\to\infty} N \cdot \int_0^1 I[(1+\Delta_N)u^{\Delta_N} \ge c_N](1+\Delta_N)u^{\Delta_N}\,\mathrm{d}u \\
&\quad = \lim_{N\to\infty} N \cdot \int_0^{1+\Delta_N} I[v \ge c_N] v^{1/\Delta_N} \Delta_N^{-1} (1+\Delta_N)^{-1/\Delta_N}\,\mathrm{d}v = 0
\end{aligned}
$$

because the set $\{v : c_N \le v \le 1+\Delta_N\}$ is empty for $N > N_0$.

(iii) Similarly, for the alternative (3.5), we have

$$
\begin{aligned}
&\sum_{k=m+1}^{N} H^2(F(z_k), G_{\Delta_N}(z_k)) \\
&\quad = n \cdot \int_0^\infty f(z)\Big[\sqrt{(1+\Delta_N)F^{\Delta_N}(z)} - 1\Big]^2 \,\mathrm{d}z
\end{aligned}
$$

$$
\begin{aligned}
&= n\int_0^\infty f(z)\frac{[(1-\Delta_N)F^{\Delta_N}(z)-1]^2}{[\sqrt{(1+\Delta_N)F^{\Delta_N}(z)}+1]^2}\,\mathrm{d}z \\
&\le n\int_0^1 [(1+\Delta_N)u^{\Delta_N}-1]^2\,\mathrm{d}u \le n\int_0^1 [\Delta_N^2+(u^{\Delta_N}-1)^2]\,\mathrm{d}u \\
&\le n\left\{\Delta_N^2+\left[\frac{u^{1+2\Delta_N}}{1+2\Delta_N}-\frac{2u^{1+\Delta_N}}{1+\Delta_N}+u\right]_0^1\right\} = n\left\{\Delta_N^2+\frac{2\Delta_N^2}{(1+2\Delta_N)(1+\Delta_n)}\right\} \\
&\le 7n\Delta_N^2 = 7\lambda_N\Delta_0^2 < \infty.
\end{aligned}
$$

Condition (A.2) is verified analogously as for the alternative (3.2). □

Acknowledgments

The authors would like to thank the Editor and two Referees for their valuable comments which helped to provide a better understanding of the whole text. They also wish to thank Pranab K. Sen, Hannu Oja and Marek Omelka for valuable discussions and M. Omelka also for his help with calculating the powers of some tests. This research was supported by the Project LC06024 of Ministry of Education, Youth and Sports of Czech Republic. J. Jurečková was also supported by the grant IAA101120801 of the Academy of Science of Czech Republic, by the Czech Republic Grant 201/09/0133 and by the research project MSM 0021620839 of the Ministry of Education, Youth and Sports of Czech Republic.

References

[1] Amrhein, P. (1995). An example of a two-sided Wilcoxon signed rank test which is not unbiased. *Ann. Inst. Statist. Math.* **47** 167–170. MR1341213

[2] Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870

[3] Bickel, P.J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Statist.* **40** 1–23. MR0256519

[4] Brown, B.M. (1982). Cramér–von Mises distributions and permutation tests. *Biometrika* **69** 619–624. MR0695207

[5] Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension: Inference based on the geometry of the data cloud. *J. Amer. Statist. Assoc.* **88** 1363–1370. MR1245371

[6] Choi, K. and Marden, J.I. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.* **92** 1581–1590. MR1615267

[7] Choi, K. and Marden, J.I. (2005). Tests of multivariate linear models using spatial concordances. *J. Nonparametr. Statist.* **17** 167–185. MR2112519

[8] Friedman, J.H. and Rafsky, L.C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

[9] Gibbons, J.D. (1964). A proposed two-sample test and its properties. *J. Roy. Statist. Soc. Ser. B* **26** 305–312. MR0174121

[10] Grose, S.D. and King, M.L. (1991). The locally unbiased two-sided Durbin–Watson test. *Econom. Lett.* **35** 401–407.
[11] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.
[12] Hájek, J., Šidák, Z. and Sen, P.K. (1999). *Theory of Rank Tests*. New York: Academic Press. MR1680991
[13] Hallin, M. and Pandaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170
[14] Hall, P. and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika* **89** 359–374. MR1913964
[15] Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. MR0947577
[16] Henze, N. and Penrose, M.D. (1999). On the multivariate runs test. *Ann. Statist.* **27** 290–298. MR1701112
[17] Hetmansperger, T.P., Möttönen, J. and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica* **8** 785–800. MR1651508
[18] Jurečková, J. (2002). L_1 derivatives, score functions and tests. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, ed.) 183–189. Basel: Birkhäuser.
[19] Jurečková, J. and Milhaud, X. (2003). Derivative in the mean of a density and statistical applications. In *Mathematical Statistics and Applications. Festschrift for Constance van Eeden* (M. Moore, C. Léger and S. Froda, eds.). *IMS Lecture Notes* **42** 217–232. Beachwood, OH: Inst. Math. Statist. MR2138294
[20] Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–42. MR0054208
[21] Liu, R. (1988). On a notion of simplicial depth. *Proc. Natl. Acad. Sci. USA* **85** 1732–1734. MR0930658
[22] Lehmann, E.L. (1997). *Testing Statistical Hypotheses*, 2nd ed. New York: Springer. MR1481711
[23] Liu, R. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400
[24] Liu, R. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489
[25] Maa, J.-F., Pearl, D.K. and Bartoszynski, R. (1996). Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Ann. Statist.* **24** 1069–1074. MR1401837
[26] Neuhaus, G. and Zhu, L.-X. (1999). Permutation tests for multivariate location problems. *J. Multivariate Anal.* **69** 167–192. MR1703370
[27] Obenchain, R.L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42** 1569–1578. MR0343463
[28] Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Austr. J. Statist.* **29** 91–100. MR0899380
[29] Oja, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. New York: Springer. MR2598854
[30] Oja, H., Möttönen, J. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. MR1439313
[31] Oja, H. and Randles, R.H. (2004). Multivariate nonparametric tests. *Statist. Science* **19** 598–605. MR2185581
[32] Oosterhoff, J. and van Zwet, W.R. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jurečková, ed.) 157–166. Dordrecht: Reidel. MR0561267
[33] Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis.* New York: Wiley. MR0298844

[34] Randles, R.H. and Peters, D. (1990). Multivariate rank tests for the two-sample location problem. *Comm. Statist. Theory Methods* **19** 4225–4238. MR1103009

[35] Rosenbaum, P.R. (2005). An exact distribution free test comparing two multivariate distributions based on adjacency. *J. Roy. Statist. Soc. Ser. B* **67** 515–530. MR2168202

[36] Savage, I.R. (1956). Contributions to the theory of rank order statistics: The two sample case. *Ann. Math. Statist.* **27** 590–615. MR0080416

[37] Schilling, M.F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. MR0860514

[38] Sugiura, N. (1965). An example of the two-sided Wilcoxon test which is not unbiased. *Ann. Inst. Statist. Math.* **17** 261–263. MR0184360

[39] Sugiura, N., Murakami, H., Lee, S.K. and Maeda, Y. (2006). Biased and unbiased two-sided Wilcoxon tests for equal sample sizes. *Ann. Inst. Statist. Math.* **58** 93–100. MR2256155

[40] Topchii, A., Tyurin, Y. and Oja, H. (2003). Inference based on the affine invariant multivariate Mann–Whitney–Wilcoxon statistic. *J. Nonparametr. Statist.* **14** 403–414. MR2017477

[41] Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proc. Intern. Congress of Mathematicians* **2** 523–531. Montréal: Canadian Mathematics Congress. MR0426989

[42] Wellner, J.A. (1979). Permutation tests for directional data. *Ann. Statist.* **7** 929–943. MR0536498

[43] Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.* **34** 2879–2896. MR2329471

Received November 2008 and revised September 2010