UNBIASEDNESS OF MULTIVARIATE TESTS

Jana Jurečková* and Jan Kalina

Charles University in Prague

1. Introduction
2. Unbiased tests
3. Rank tests
4. Structure of the rank tests
5. Asymptotic efficiencies
6. References

Introduction

Many tests are surprisingly biased when based on finite numbers of observations. Lehmann (1959) showed that two-sample Wilcoxon test is unbiased against one-sided alternatives, and put a question whether it is also unbiased against two-sided alternatives. This question was answered by Sugiura (1965), who showed that generally this is not the case, not even with equal sample sizes (see also Sugiura et al. 2006). Amrhein (1995) demonstrated the same phenomenon for the one-sample Wilcoxon test. The question of unbiasedness is more serious in two-sample multivariate models. The tests, based on some ordered data clouds, are generally not unbiased against two-sample alternatives, while one-sided alternatives in data clouds have a difficult interpretation in original observations.

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.
The tests described in the literature are typically consistent against distant alternatives, and some of them are affine invariant. The authors of tests mostly derive their asymptotic null distributions, and sometimes the asymptotic powers under contiguous alternatives. They illustrate the powers on simulated data, often normally distributed, and compare the power with that of the Hotelling T^2 test. However, the finite sample unbiasedness of many tests is still an open question.

Let random vector $\mathbf{X} = (X_1, \ldots, X_n)$ have 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 positive definite Fisher information matrix and bounded third derivative in components of $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. Our goal: To test $\mathbf{H}_0: \ \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against alternative $\mathbf{K}: \ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ by test Φ of size α, i.e. $\mathbb{E}_{\boldsymbol{\theta}_0}[\Phi(\mathbf{X})] = \alpha$. Jur. & Milhaud (2003) derived the following expansion of the power function of Φ around $\boldsymbol{\theta}_0$:

$$\mathbb{E}_{\boldsymbol{\theta}}\Phi(\mathbf{X}) = \alpha + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbb{E}_{\boldsymbol{\theta}_0}\left\{\Phi(\mathbf{X})\frac{(\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}_0))}{f(\mathbf{X}, \boldsymbol{\theta}_0)}\right\} \tag{1}$$

$$+\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbb{E}_{\boldsymbol{\theta}_0}\left\{\Phi(\mathbf{X})\frac{[\ddot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}_0)]}{f(\mathbf{X}, \boldsymbol{\theta}_0)}\right\}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathcal{O}\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^3\right),$$

where

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

The test Φ is locally unbiased if the second term on the right-hand side of (1) is ≥ 0. For scalar θ, there always exists an unbiased test against one-sided alternative $\mathbf{K}: \ \theta > \theta_0$. For a vector $\boldsymbol{\theta}$, the local (two-sided) unbiasedness of Φ is guaranteed only when

$$\mathbb{E}_{\theta_0}\left\{\Phi(\mathbf{X})\frac{(\dot{\mathbf{f}}(\mathbf{X},\boldsymbol{\theta}_0))}{f(\mathbf{X},\boldsymbol{\theta}_0)}\right\} = \mathbf{0}. \qquad (2)$$

But (2) is generally true only for f satisfying special conditions; otherwise the second term in (1) can be < 0 for some $\boldsymbol{\theta}$, leading to the power of Φ less than α. Grose and King (1991) imposed condition (2) when they constructed a locally unbiased two-sided version of the Durbin-Watson test.

Affine invariance

Let $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$ and $\mathcal{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ be independent samples from continuous distribution functions $F^{(p)}$ and $G^{(p)}$, with means and dispersion matrices $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$. We wish to test the hypothesis $\mathbf{H}_0 : F^{(p)} \equiv G^{(p)}$ against 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. Denote $(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$ the pooled sample. $\mathbf{H}_0$ and $\mathbf{H}_1$ are invariant under affine transformations

$$\mathcal{G} : \{\mathbf{Z} \to \mathbf{a} + \mathbf{B}\mathbf{Z}\}, \quad \mathbf{a} \in \mathbb{R}^p, \ \mathbf{B} \text{ nonsingular } \ p \times p.$$

The tests invariant to $g \in \mathcal{G}$ should depend on data only by means of a *maximal invariant* of $\mathcal{G}$.

Obenchain (1971) showed that the maximal invariant with respect to $\mathcal{G}$ is

$$\mathbf{T}(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = \left[(\mathbf{Z}_i - \bar{\mathbf{Z}}_N)^\top \mathbf{V}_N^{-1} (\mathbf{Z}_j - \bar{\mathbf{Z}}_N)\right]_{i,j=1}^N$$

where $\bar{\mathbf{Z}}_N = \frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i, \ \mathbf{V}_N = \sum_{i=1}^N (\mathbf{Z}_i - \bar{\mathbf{Z}}_N)(\mathbf{Z}_i - \bar{\mathbf{Z}}_N)^\top$.
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 $\left[\mathbf{Z}_1 - \bar{\mathbf{Z}}_N, \ldots, \mathbf{Z}_N - \bar{\mathbf{Z}}_N\right]$. Under $\mathbf{a} \equiv \mathbf{0}$, the maximal invariant of the group $\mathcal{G}_0 : \{\mathbf{Z} \to \mathbf{BZ}\}$ is

$$\mathbf{T}_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = \left[\mathbf{Z}_i^\top (\mathbf{V}_N^0)^{-1} \mathbf{Z}_j\right]_{i,j=1}^N, \ \mathbf{V}_N^0 = \sum_{i=1}^N \mathbf{Z}_i \mathbf{Z}_i^\top.$$

Moreover, one of the maximal invariants with respect to the *group of shifts in location* $\mathcal{G}_1 : \mathbf{Z} \longrightarrow \mathbf{Z} + \mathbf{a}, \ \mathbf{a} \in \mathbb{R}^p$ 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 two-sample Hotelling T^2 test based on the criterion $\mathcal{T}_{mn}^2 = (\bar{\mathbf{X}}_m - \bar{\mathbf{Y}}_n)^\top \mathbf{V}_N^{-1} (\bar{\mathbf{X}}_m - \bar{\mathbf{Y}}_n)$ is invariant with respect to $\mathcal{G}$ and optimal unbiased against two-sample normal alternatives with $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. However, its finite sample unbiasedness is not guaranteed under non-normal underlying distributions.

The multivariate two-sample tests, based on scalar geometric structures of data, are unbiased to alternatives, one-sided in ordering these structures, or even to two-sided alternatives, if distribution of these structures is symmetric under $\mathbf{H}_0$. Hence, we should look for geometric structures for which the one-sided alternatives are meaningful, or whose distribution is symmetric under $\mathbf{H}_0$. One-sided alternatives have a sense for distances, definitely for Euclidean distances. But we have a problem of invariance of tests, based upon them, to the change of origin.

An unbiased test, invariant to $\mathcal{G}$, should be a function of Mahalanobis distances. Perhaps some transform can guarantee their symmetry under $\mathbf{H}_0$. A suitable rank test, based on the ranks of Mahalanobis distances, should be still investigated. Their rank properties are not trivial. Optimally, we would like to construct tests which are

(i) distribution free under $\mathbf{H}_0$,
(ii) affine invariant with respect to changes of coordinate system,
(iii) consistent against any fixed alternative,
(iv) 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. We can construct rank tests with good properties,enjoying properties (i), (iii) and (iv), but invariant only to $\mathcal{G}_1$, not to the change of origin.

Liu and Singh rank sum test

Liu and Singh (1993) proposed a two-sample test of Wilcoxon type, based on the ranks of depths of the data. Its asymptotic distributions under the hypothesis and under general alternative distributions F, G of depths was derived by Zuo and He (2006). Let $D(\mathbf{y}; H)$ denote a depth function of a distribution H evaluated at point $\mathbf{y} \in \mathbb{R}^p$. Let a *"quality index"* be defined as

$$Q(F^{(p)}, G^{(p)}) = \int R(\mathbf{y}; F^{(p)}) dG^{(p)}(\mathbf{y})$$
$$= \mathbb{P}\left\{ D(\mathbf{X}; F^{(p)}) \leq D(\mathbf{Y}; F^{(p)}) \middle| \mathbf{X} \sim F^{(p)}, \mathbf{Y} \sim G^{(p)} \right\}$$

where $R(\mathbf{y}; F^{(p)}) = \mathbb{P}_F \left\{ D(\mathbf{X}; F^{(p)}) \leq D(\mathbf{y}; F^{(p)} \right\}, \ \mathbf{y} \in \mathbb{R}^p$. If $D(\mathbf{X}; F^{(p)})$ has a continuous distribution, then $Q(F^{(p)}, F^{(p)}) = \frac{1}{2}$.

Liu and Singh tested hypothesis $Q(F^{(p)}, G^{(p)}) = \frac{1}{2}$ against alternative $Q(F^{(p)}, G^{(p)}) \neq \frac{1}{2}$, using the Wilcoxon criterion based on empirical distribution functions F_m, G_n of samples of sizes m, n :

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

If distribution of depths is symmetric under $F^{(p)} \equiv G^{(p)}$, then the test rejecting provided $|Q(F_m, G_n) - \frac{1}{2}| \geq C_{\alpha/2}$ is locally unbiased against $Q(F^{(p)}, G^{(p)}) \neq \frac{1}{2}$. Under general distribution of depths, the test with the critical region $Q(F_m, G_n) - \frac{1}{2} > C_\alpha$ is unbiased against 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 distributions $F^{(p)}, G^{(p)}$ of original observations $\mathbf{X}$ and $\mathbf{Y}$. Generally, the test is not finite-sample unbiased against $F \neq G$, not even locally. The unbiasedness can be guaranteed only in some cases, e.g. if the hypothetical distribution of depths is symmetric.

Rank tests

We propose three classes of multivariate two-sample tests, based on the ranks of suitable distances of observations. The distances are either those of observations from origin, or inter-point distances. The natural (one-sided) alternatives state that distances of the second sample are stochastically larger than those of the first sample; hence the tests are unbiased against such alternatives. The proposed rank tests are distribution free under the hypothesis but also under some alternatives (of the Lehmann type), hence we can find the critical value and the power. They are consistent against general alternatives, possessing properties (i), (iii) and (iv).

Let $\mathcal{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = (\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ for $N = m + n$ be the pooled sample. Select a nonnegative distance $L = L(\cdot, \cdot)$ in $\mathbb{R}^p$ and consider the matrix $\mathbb{L}_N = [\ell_{ik}]_{i,k=1}^N$, $\ell_{ik} = L(\mathbf{Z}_i, \mathbf{Z}_k)$. Rank tests can be based on $\mathbb{L}_N$ in three ways:

(i) Simple rank test is based on the vector $(\tilde{\ell}_1, \ldots, \tilde{\ell}_N)$, $\tilde{\ell}_k = L(\mathbf{0}, \mathbf{Z}_k)$, $k = 1, \ldots, N$ of distances from the origin. Then vectors $(\tilde{\ell}_1, \ldots, \tilde{\ell}_m)$ and $(\tilde{\ell}_{m+1}, \ldots, \tilde{\ell}_N)$ are random samples from distribution functions F and G (say), respectively, assumed being absolutely continuous. Under $\mathbf{H}_0 : \; F^{(p)} \equiv G^{(p)}$, the F and G coincide. The vector $\widetilde{\mathcal{R}} = (\tilde{R}_1, \ldots, \tilde{R}_N)$ of ranks of $\{\tilde{\ell}_k, \; k = 1, \ldots, N\}$ is then uniformly distributed on the permutations of numbers $1, \ldots, N$. Every two-sample rank test is invariant to increasing continuous functions of $(\tilde{\ell}_1, \ldots, \tilde{\ell}_N)$. However, it is not invariant to the groups of transformations $\mathcal{G}$ or $\mathcal{G}_1$, even if $\tilde{\ell}_k = \|\mathbf{Z}_k\|$ is Euclidean distance.

Such rank test is based on the linear rank statistic $S_N = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_{ik})$. The test is distribution-free under $\mathbf{H}_0$, and its properties are well-known.

(ii) Conditional rank test, invariant to $\mathcal{G}_1$. Assuming that $m > p$, fix a base $(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_p}) = \mathbb{X}_p$ of $\{\mathbf{X}_i,\ 1 \leq i \leq m\}$. Consider the set of $(m+n-p) \times p$ distances $\{\ell^*_{i_j,k} = L(\mathbf{X}_{i_j}, \mathbf{Z}_k)$, $k = 1, \ldots, N,\ k \neq i_1, \ldots, i_p\},\ j = 1, \ldots, p$. Then, for a fixed $i_j,\ 1 \leq j \leq p$, and conditionally given $\mathbb{X}_p$, the vectors $\{\ell^*_{i_j,k},\ k = 1, \ldots, m,\ k \neq i_1, \ldots, i_p\}$ and $\{\ell^*_{i_j,k},\ k = m+1, \ldots, N\}$ are random samples from distribution functions $F(z|\mathbb{X}_p) = F$ and $G(z|\mathbb{X}_p) = G$, (say), assumed being absolutely continuous.

We can consider two-sample linear (conditional) rank test based on the linear rank statistic $S^*_{i_j,N} = N^{-1/2}\sum_{k=m+1}^{N} a_N(R_{i_j,k})$ where $\mathcal{R}_{i_j} = (R_{i_j,k},\ k=1,\ldots,N,\ k \neq i_1,\ldots,i_p)$ are the ranks of $\ell^*_{i_j k},\ k=1,\ldots,N,\ k\neq i_j$. Particularly, if $L(\mathbf{X}_{i_j},\mathbf{Z}_k) = \|\mathbf{X}_{i_j} - \mathbf{Z}_k\|$, $k=1,\ldots,N,\ k\neq i_1,\ldots,i_p$, where $\|\cdot\|$ is the Euclidean distance, then the test based on their ranks will be invariant to $\mathcal{G}_1$, but not to $\mathcal{G},\ \mathcal{G}_0$. The criteria $S^*_{i_j,N}$ are equally distributed for $j=1,\ldots,p$, under the hypothesis and under the alternatives, and conditionally independent, given $\mathbb{X}_p$. The simplest combination is a randomization of $S^*_{i_1,N},\ldots,S^*_{i_p,N}$, leading to the criterion $\widetilde{S}^{(N)}$ such that

$$\mathbb{P}(\widetilde{S}^{(N)} = S^*_{i_j,N}) = \frac{1}{p}, \quad j=1,\ldots,p$$

with the randomization independent of the set of observations $\mathcal{Z}$.

Remark

The Mahalanobis distances

$$\mathbf{Z}_k^\top (\mathbf{V}_N^0)^{-1} \mathbf{Z}_k, \;\; k = 1, \ldots, N \tag{3}$$

$$(\mathbf{X}_i - \mathbf{Z}_k)^\top \mathbf{V}_N^{-1} (\mathbf{X}_i - \mathbf{Z}_k) \;\; \text{or} \;\; (\mathbf{X}_i - \mathbf{Z}_k)^\top (\mathbf{V}_N^0)^{-1} (\mathbf{X}_i - \mathbf{Z}_k), \tag{4}$$

$k \neq i$, 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 distribution of observations (is distribution free). Moreover, (4) are invariant with respect $\mathcal{G}$ and $\mathcal{G}_0$, while (3) are invariant only with respect $\mathcal{G}_0$. The invariant tests based on ranks of (3) or (4) will be a subject of a further study. Their structure is more complex than that of tests based on simple distances.

Let $\mathbf{X} = (X_1, \ldots, X_m)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be two independent samples from distributions F, G. Consider the rank test with the criterion $S_N = N^{-1/2} \sum_{k=m+1}^{N} a_N(R_i)$ where $R_1, \ldots, R_N$ are the ranks of the pooled sample $\mathbf{Z} = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$. The values $X_i, \ Y_j$ are e.g. the distances of multivariate observations, either from a fixed point, or the inter-point distances considered conditionally given the original component. We want to test $\mathbf{H}_0: \ F \equiv G$ against a general alternative with the $(m+n)$-dimensional distribution function of the form

$$\mathbf{K}: \prod_{k=1}^{m} G_{\Delta}^{(1)}(z_k) \prod_{k=m+1}^{N} G_{\Delta}^{(2)}(z_k). \tag{5}$$

Wilcoxon test

Lehmann (1953) showed that the Wilcoxon test is the locally most powerful rank test of $\mathbf{H}_0$ against the class of alternatives (5) with

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

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

with $0 < \Delta < 1$. Then Y's are stochastically larger than 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

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

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

Psi-test

Gibbons (1964) proved that the *Psi-test* with the scores

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

with score function $\varphi(u) = \ln u - \ln(1-u))$, $0 < u < 1$ is the locally most powerful rank test of $\mathbf{H}_0$ against (5) with

$$G_\Delta^{(1)}(z) = 1 - (1 - F(z))^{1+\Delta},$$

$$G_\Delta^{(2)}(z) = (F(z))^{1+\Delta}, \ \Delta > 0, \ z \geq 0.$$

$G_\Delta^{(1)}$ is stochastically smaller than $G_\Delta^{(2)}$ for $\Delta > 0$. The Kolmogorov distance of $G_\Delta^{(1)}$ and $G_\Delta^{(2)}$ is maximized at $z = F^{-1}(\frac{1}{2})$.

Savage test

Savage (1956) proved that the test with the critical region $\sum_{i=1}^{n}\sum_{j=R_{m+i}}^{N}\frac{1}{j}\leq C_\alpha$ and with the score-generating function $\varphi(u)=1+\ln u,\ 0<u<1$ is the locally most powerful rank test of $\mathbf{H}_0$ against the class of alternatives (5) with

$$G_\Delta^{(1)}(z)=F(z),\quad G_\Delta^{(2)}(z)=F^{1+\Delta}(z),\ z\geq 0,\ \Delta>0.$$

Again, Y's are stochastically larger than X's, and the Kolmogorov distance of $G_\Delta^{(1)}$ and $G_\Delta^{(2)}$ is maximized at $z=F^{-1}\left((1+\Delta)^{-1/\Delta}\right)$. If F is increasing, then the tests are distribution free under the hypothesis as well as under the above alternatives.

Efficiencies

The above alternatives are contiguous with respect to sequence $\{\prod_{i=1}^{N} F(z_i)\}$, for $\Delta_N = N^{-1/2}\Delta_0$ with $0 < \Delta_0 < \infty$ fixed. Hence we can evaluate local asymptotic powers of the tests. Table 1 gives relative asymptotic efficiencies of Wilcoxon, Psi and Savage tests with respect to their locally most powerful rank tests. For information, we add van der Waerden and median tests.

alternative/test	Wilcoxon	Psi	Savage	Waerden	Median
(1)	1.000	0.912	0.750	0.955	0.750
(2)	0.912	1.000	0.882	0.992	0.584
(3)	0.750	0.822	1.000	0.816	0.480

Table 1. Relative asymptotic efficiencies under the above alternatives

Hotelling vs. Wilcoxon

Empirical powers of Hotelling T^2 and Wilcoxon tests are compared under bivariate normal and Cauchy distributions; Wilcoxon is based on the ranks of the euclidean interpoint distances. Sample sizes are $m = n = 10,\ 100,\ 1000$, and the simulations are based on 10 000 replications. Table 2 provides empirical powers for two bivariate normal samples. The first sample has always $\mathcal{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ distribution with $\boldsymbol{\mu}_1 = (0,0)^\top$ and $\boldsymbol{\Sigma}_1 = \text{Diag}\{1,1\}$, while the second sample has $\mathcal{N}_2(\boldsymbol{\mu}_2, \mathbf{S})$ with various parameters. Table 3 presents empirical powers for two samples from the bivariate Cauchy distributions. The first sample $\mathbf{X}$ has 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, independent on $\mathbf{X}$, transformed to $\mathbf{Y} = \boldsymbol{\mu} + \sigma\mathbf{Y}^*$ for certain shifts and scales.

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
$\mathrm{Diag}\{1,1\}$	W	0.0457	0.0505	0.0487
$\boldsymbol{\mu}_2 = (0.2,0.2)^T$	H	0.0771	0.4115	1.0000
$\mathrm{Diag}\{1,1\}$	W	0.0520	0.1715	0.6458
$\boldsymbol{\mu}_2 = (0.5,0.5)^T$	H	0.2318	0.9962	1.0000
$\mathrm{Diag}\{1,1\}$	W	0.1085	0.5701	0.8617
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0659	0.0561	0.0452
$\mathrm{Diag}\{0.1,0.1\}$	W	0.7994	0.9998	1.0000
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0653	0.0456	0.0530
$\mathrm{Diag}\{0.2,0.2\}$	W	0.4851	0.9932	1.0000
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0521	0.0521	0.0463
$\mathrm{Diag}\{0.5,0.5\}$	W	0.1182	0.7034	0.9968
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0531	0.0530	0.0514
$\mathrm{Diag}\{1.5,1.5\}$	W	0.0656	0.2881	0.8525
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0552	0.0518	0.0508
$\mathrm{Diag}\{2,2\}$	W	0.0999	0.5395	0.9670
$\boldsymbol{\mu}_2 = (0,0)^T$	H	0.0572	0.0546	0.0521
$diag\{1.0,0.2\}$	W	0.1029	0.6568	0.9936

Table 2. Powers of two-sample Hotelling T^2 test (H) and of two-sample Wilcoxon test (W), $\alpha = 0.05$, bivariate normal

Second sample	Test	$m, n = 10$	$m, n = 25$	$m, n = 100$	$m, n = 1000$
$(0,0)^T$	H	0.0191	0.0156	0.0171	0.0217
$\sigma = 1$	W	0.0450	0.0478	0.0510	0.0442
$(0.2,0.2)^T$	H	0.0227	0.0232	0.0227	0.0174
$\sigma = 1$	W	0.0468	0.0536	0.0874	0.3925
$(0.5,0.5)^T$	H	0.0408	0.0404	0.0414	0.0361
$\sigma = 1$	W	0.0664	0.1115	0.2937	0.7470
$(1,1)^T$	H	0.1038	0.1193	0.1260	0.1226
$\sigma = 1$	W	0.1219	0.2710	0.6235	0.8893
$(5,5)^T$	H	0.7387	0.7535	0.7683	0.7772
$\sigma = 1$	W	0.7574	0.9441	0.9782	0.9944
$(0,0)^T$	H	0.0200	0.0171	0.0193	0.0103
$\sigma = 1.5$	W	0.0664	0.1207	0.3419	0.8428
$(0,0)^T$	H	0.0207	0.0168	0.0182	0.0172
$\sigma = 2$	W	0.1082	0.2439	0.6123	0.9135
$(0.2,0.2)^T$	H	0.0189	0.0201	0.0196	0.0240
$\sigma = 1.5$	W	0.0710	0.1297	0.3495	0.8249
$(1,1)^T$	H	0.0741	0.0814	0.0865	0.0943
$\sigma = 1.5$	W	0.1088	0.2188	0.4925	0.8462

Table 3. Powers of two-sample Hotelling T^2 (H) and two-sample Wilcoxon (W) tests, $\alpha = 0.05$, *bivariate Cauchy*

Conclusion

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 another scalar characteristics which are symmetrically distributed under the hypothesis, then the tests are locally unbiased also against the two sample alternatives. The Lehmann alternatives reflect well the practical situations.

References

- Amrhein, P. (1995). An example of a two-sided Wilcoxon signed rank test which is not unbiased. *Ann. Inst. Statist. Math.* **47**, 167–170.
- Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *J. Multiv. Anal.* **88**, 190–206.
- Bickel, P. J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Statist.* **40**, 1–23.
- Brown, B. M. (1982) Cramér-von Mises distributions and permutation tests. *Biometrika* **69**, 619–624.
- 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.

- Choi, K. and Marden, J. I. (2005). Tests of multivariate linear models using spatial concordances. *J. Nonparametr. Statist.* **17**, 167–185.
- Friedman, J. H., and Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7**, 697–717.
- Gibbons, J. D. (1964). A proposed two-sample test and its properties. *J. R. Statist. Soc. B* **26**, 305–312.
- Grose, S. D. and King, M. L. (1991). The locally unbiased two-sided Durbin-Watson test. *Economics Letters* **35**, 401–407.
- Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academia, Prague & Academic Press, New York.
- Hájek, J., Šidák, Z., and Sen, P. K. (1999). *Theory of Rank Tests*, Academic Press, New York.

- Hallin, M. and Pandaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30**, 1103–1133.
- Hall, P., and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika* 89, 359–374.
- Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16**, 772–783.
- Henze, N., and Penrose, M. D. (1999). On the multivariate runs test, *Ann. Statist.* **27**, 290–298.
- Hetmansperger, T. P., Möttönen, J., and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statistica Sinica* **8**, 785–800.

Jurečková, J. (2002). L_1 derivatives, score functions and tests. *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, ed.), pp. 183–189. Birkhäuser, Basel.

Jurečková, J., and Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli* 18, 229-251.

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, S. Froda, eds.) *IMS Lecture Notes* **42**, pp. 217–232.

Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Statist.* **24**, 23–42.

Lehmann, E.L. (1997). *Testing Statistical Hypotheses* (2nd edition). Springer, New York.

Liu, R. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18**, 405–414.

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

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.

Neuhaus, G., and Zhu, L.-X. (1999). Permutation tests for multivariate location problems. *J. Multiv. Anal.* **69**, 167–192.

Obenchain, R. L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42**, 1569–1578

- Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Austr. J. Statist.* **29**, 91–100.
- Oja, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. Lecture Notes in Statistics 199, Springer.
- Oja, H., Möttönen, J. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25**, 542–552.
- Oja, H. and Randles, R. H. (2004). Multivariate nonparametric tests. *Statistical Science* **19**, 598-605.
- 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.). Prague, Academia & Dordrecht, Reidel, pp. 157–166.

- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* J. Wiley, New York.
- Randles, R. H. and Peters, D. (1990). Multivariate rank tests for the two-sample location problem. *Comm. Stat. Theory Methods* **19**, 4225–4238.
- Rosenbaum, P. R. (2005). An exact distribution free test comparing two multivariate distributions based on adjacency. *J. R. Statist. Soc. B* **67**, 515–530.
- Savage, I. R. (1956). Contributions to the theory of rank order statistics - The two sample case. *Ann. Math. Statist.* **27**, 590–615.
- Schilling, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Ass.* **81**, 799–806.
- Sugiura, N. (1965). An example of the two-sided Wilcoxon test which is not unbiased. *Ann. Inst. Statist. Math.* **17**, 261–263.

- 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.
- 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.
- Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proc. Intern. Congress of Mathematicians* 2, Montréal: Canadian Mathematics Congress, pp. 523–531.
- Wellner, J. A. (1979). Permutation tests for directional data. *Ann. Statist.* **7**, 929–943.
- 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.