Rank scores tests of multivariate independence

S. Taskinen, A. Kankainen, and H. Oja

Abstract. New rank scores test statistics are proposed for testing whether two random vectors are independent. The tests are asymptotically distribution-free for elliptically symmetric marginal distributions. Recently, Gieser and Randles (1997), Taskinen, Kankainen and Oja (2003) and Taskinen, Oja and Randles (2005) introduced and discussed different multivariate extensions of the quadrant test, Kendall's tau and Spearman's rho statistics. In this paper, standardized multivariate spatial signs and the (univariate) ranks of the Mahalanobis-type distances of the observations from the origin are combined to construct rank scores tests of independence. The limiting distributions of the test statistics are derived under the null hypothesis as well as under contiguous sequences of alternatives. Three different choices of the score functions, namely the sign scores, the Wilcoxon scores and the van der Waerden scores, are discussed in greater detail. The small sample and limiting efficiencies of the test procedures are compared and the robustness properties are illustrated by an example. It is remarkable that, in the multinormal case, the limiting Pitman efficiency of the van der Waerden scores test equals to that of the classical parametric Wilks' test.

1. Introduction

Let $\boldsymbol{x}_i^T = (\boldsymbol{x}_i^{(1)^T}, \boldsymbol{x}_i^{(2)^T})$ for $i = 1, \ldots, n$ be a random sample of vector pairs, where $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ are p- and q-dimensional continuous random vectors. We wish to test the null hypothesis

$$H_0: \quad \boldsymbol{x}_i^{(1)} \text{ and } \boldsymbol{x}_i^{(2)} \text{ are independent.}$$

The classical parametric test due to Wilks (1935) is based on the partitioned sample covariance matrix S and is defined as

$$W_n = \frac{|S|}{|S_{11}||S_{22}|}.$$

Puri and Sen (1971) introduced a nonparametric analogue to Wilks' test where the data vectors are replaced by the vectors of their componentwise ranks. Gieser and Randles (1997) and Taskinen et al. (2003) proposed invariant extensions of the univariate quadrant test of Blomqvist (1950). The former test procedure is based on interdirection counts and the latter on standardized spatial signs. If the marginal distributions of $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ are elliptic, these two tests are asymptotically equivalent. Later Taskinen et al. (2005) proposed multivariate invariant extensions of Kendall's tau and Spearman's rho.

Our plan is as follows. In Section 2, we explain the test constructions starting with standardized spatial signs and ranks of the lengths of the standardized vectors. The test statistics for multivariate dependence are then introduced. Special choices of the score functions then yield the sign test, the Wilcoxon scores test and the van der Waerden scores test. In Section 3, the limiting distribution of the test statistic is derived under the null hypothesis and under interesting sequences of contiguous alternatives. The finite-sample and limiting efficiencies of the new procedures are then compared to that of the classical Wilks' test in Section 4, and the robustness properties are illustrated by an example in the final Section 5. The proofs are postponed to Appendix I.

2. The rank scores test statistics

2.1. Spatial signs and ranks of the distances from the origin

Consider a random sample $\boldsymbol{x}_1, \dots \boldsymbol{x}_n$ from a k-variate distribution. The spatial sign of vector $\boldsymbol{x}$ is defined as

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{-1}\boldsymbol{x}, & \boldsymbol{x} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{cases}$$

where $\|\boldsymbol{x}\| = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}$ is the (Euclidean) length of the vector $\boldsymbol{x}$. The spatial signs $\boldsymbol{S}(\boldsymbol{x}_i)$ and ranks $\mathrm{rank}(\|\boldsymbol{x}_i\|)$ of the distances from the origin are not invariant under affine transformations to the data vectors, however. In order to construct invariant test statistics, the data points have to be standardized before spatial signs and ranks are formed. For the standardization we need affine equivariant $\sqrt{n}$-consistent location vector and scatter matrix estimates, $\widehat{\boldsymbol{\mu}}$ and $\widehat{C}$. The transformed data points are then given as $\boldsymbol{z}_i = \widehat{C}^{-1/2}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})$, $i = 1, \dots, n$.

The vectors $\widehat{\boldsymbol{u}}_i = \boldsymbol{S}(\boldsymbol{z}_i)$, $i = 1, \dots, n$, are called **standardized spatial sign vectors**. Standardized sign vectors are affine invariant in the sense that if $\widehat{\boldsymbol{u}}_i^*$ are calculated from $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$, $i = 1, \dots, n$, with a nonsingular $k \times k$ matrix A and k-vector $\boldsymbol{b}$, then $\widehat{\boldsymbol{u}}_i^* = P\widehat{\boldsymbol{u}}_i$, $i = 1, \dots, n$, for some orthogonal P. See e.g. Taskinen et al. (2005). The ranks $\widehat{R}_i = \mathrm{rank}(\|\boldsymbol{z}_i\|)$ are naturally affine invariant (in the usual sense). Note that, in the standardization, the scatter matrix estimate $\widehat{C}$ may be replaced by a $\sqrt{n}$-consistent affine equivariant shape matrix estimate $\widehat{V}$ as only the directions and ranks of distances are used in the analysis. For the shape matrices, see Ollila et al. (2003). Note also that, if the standardization is

done using such location vector and scatter (or shape) matrix estimates that do not require any moment assumptions of the underlying data (e.g. Tyler's shape matrix and the transformation retransformation spatial median in Hettmansperger and Randles, 2002), then the resulting test procedures are valid without any moment assumptions.

2.2. New test statistics

Our test statistic for testing the null hypothesis of independence is obtained as follows. For $i = 1, \ldots, n$, write $\widehat{\boldsymbol{u}}_i^{(1)} = ||\boldsymbol{z}_i^{(1)}||^{-1} \boldsymbol{z}_i^{(1)}$ for p-dimensional standardized sign vectors based on the first components $\boldsymbol{x}_i^{(1)}$ and let $\widehat{R}_i^{(1)}$ denote the rank of $\widehat{r}_i^{(1)} = ||\boldsymbol{z}_i^{(1)}||$ among $\widehat{r}_1^{(1)}, \ldots, \widehat{r}_n^{(1)}$. For the second random vector, write similarly $\widehat{\boldsymbol{u}}_i^{(2)}$ for q-dimensional standardized sign vectors based on $\boldsymbol{x}_i^{(2)}$ and let $\widehat{r}_i^{(2)}$ and $\widehat{R}_i^{(2)}$ be constructed as before. The test statistic is then as follows.

Definition 2.1. Let $a : (0,1) \to \mathbb{R}$ and $b : (0,1) \to \mathbb{R}$ be continuous, monotone and square integrable score functions and write

$$\widehat{H} = \text{ave}_i \left\{ a\left(\frac{\widehat{R}_i^{(1)}}{n+1}\right) b\left(\frac{\widehat{R}_i^{(2)}}{n+1}\right) \widehat{\boldsymbol{u}}_i^{(1)} \widehat{\boldsymbol{u}}_i^{(2)^T} \right\}.$$

The rank test statistic for testing H_0 is then

$$T_n = \frac{npq}{\sigma_a^2 \sigma_b^2} ||\widehat{H}||^2,$$

where $||\widehat{H}||^2 = Tr(\widehat{H}^T \widehat{H})$, $\sigma_a^2 = E[a^2(U)]$ and $\sigma_b^2 = E[b^2(U)]$ with U uniformly distributed on $(0,1)$.

Note that since standardized sign vectors and ranks are invariant with respect to the group of affine transformations, the invariance of T_n easily follows. As score functions, one may use optimal location score functions. See Hallin and Paindaveine (2002), for that. In the following, some choices of the score functions and resulting test statistics are given.

Definition 2.2. For $a(u) = 1$ and $b(u) = 1$, the **sign test** of independence (Taskinen et al., 2003) with test statistic

$$T_{0n} = npq \, ||\text{ave}_i \{\widehat{\boldsymbol{u}}_i^{(1)} \widehat{\boldsymbol{u}}_i^{(2)^T}\}||^2$$

is obtained. For $a(u) = u$ and $b(u) = u$, one gets the **Wilcoxon (scores) test** of independence with test statistic

$$T_{1n} = \frac{9npq}{(n+1)^4} ||\text{ave}_i \{\widehat{R}_i^{(1)} \widehat{R}_i^{(2)} \widehat{\boldsymbol{u}}_i^{(1)} \widehat{\boldsymbol{u}}_i^{(2)^T}\}||^2.$$

Finally the choices $a(u) = [\Psi_p^{-1}(u)]^{1/2}$ and $b(u) = [\Psi_q^{-1}(u)]^{1/2}$, where Ψ_k is a cdf of chi-square distribution with k degrees of freedom, yield the **van der Waerden**

(**scores**) **test** of independence with test statistic

$$T_{2n} = n \left\| \mathrm{ave}_i \left\{ \left[\Psi_p^{-1} \left(\frac{\widehat{R}_i^{(1)}}{n+1} \right) \right]^{1/2} \left[\Psi_q^{-1} \left(\frac{\widehat{R}_i^{(2)}}{n+1} \right) \right]^{1/2} \widehat{u}_i^{(1)} \widehat{u}_i^{(2)^T} \right\} \right\|^2.$$

3. Limiting distributions

In order to derive the limiting distribution of T_n, we assume that the marginal distributions of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are elliptically symmetric. The marginal density functions are then of the form

$$f(\boldsymbol{x}) = |\Sigma|^{-1/2} f_0(\Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})),$$

where Σ is a positive definite symmetric matrix and $f_0(\boldsymbol{z}) = \exp\{-\rho(||\boldsymbol{z}||)\}$ with $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$. Note that if $r = ||\boldsymbol{z}||$ and $\boldsymbol{u} = \boldsymbol{z}/r$, then r and $\boldsymbol{u}$ are independent. In the following we denote the cdf of $r^{(1)}$ as G_1 and the cdf of $r^{(2)}$ as G_2.

To establish a limiting distribution of our test statistic under the null hypothesis, we need the following lemma.

Lemma 3.1. *Let*

$$H = \mathrm{ave}_i \{ a(G_1(r_i^{(1)})) b(G_2(r_i^{(2)})) \boldsymbol{u}_i^{(1)} \boldsymbol{u}_i^{(2)^T} \},$$

then $\sqrt{n}(\widehat{H} - H) \to_p 0$.

Now the limiting distribution can be found easily.

Theorem 3.2. *Under H_0 and for elliptically distributed $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, the limiting distribution of T_n is a chi-square distribution with pq degrees of freedom.*

Next we derive the limiting distribution of T_n under alternative sequences similar to those used in Gieser and Randles (1997). As T_n is affine invariant, we restrict to the spherical case only. See Appendix II, for a discussion on the alternative sequences. Let thus $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ be independent with spherical marginal densities $\exp\{-\rho_1(||\boldsymbol{x}^{(1)}||)\}$ and $\exp\{-\rho_2(||\boldsymbol{x}^{(2)}||)\}$, respectively, and write

$$(3.1) \qquad \begin{pmatrix} \boldsymbol{y}_i^{(1)} \\ \boldsymbol{y}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (1-\Delta)I_p & \Delta M_1 \\ \Delta M_2 & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix},$$

where $\Delta = \delta/\sqrt{n}$. If T_n^* is calculated from transformed observations in (3.1), we get

Theorem 3.3. *Under general assumptions (stated in the Appendix), the limiting distribution of T_n^* is a noncentral chi-square distribution with pq degrees of freedom and noncentrality parameter*

$$\frac{\delta^2}{pq\sigma_a^2\sigma_b^2} ||c_1 M_1 + c_2 M_2^T||^2,$$

where

$$c_1 = E[a(G_1(r_i^{(1)}))\psi_1(r_i^{(1)})] E[b(G_2(r_i^{(2)}))r_i^{(2)}]$$

and

$$c_2 = E[b(G_2(r_i^{(2)}))\psi_2(r_i^{(2)})]E[a(G_1(r_i^{(1)}))r_i^{(1)}],$$

with optimal location score functions $\psi_1(r_i^{(1)}) = \rho_1'(r_i^{(1)})$ *and* $\psi_2(r_i^{(2)}) = \rho_2'(r_i^{(2)})$.

4. Limiting and finite-sample efficiencies

4.1. Limiting Pitman efficiencies

In this section we consider the sign, Wilcoxon and van der Waerden tests of independence: We compare the limiting and finite-sample efficiencies of the new tests to those of the Wilks' likelihood ratio test W_n. The comparisons are made in the multivariate normal distribution, t distribution and contaminated normal distribution cases. Since $-n \log W_n$ has, under the alternative sequences, a limiting noncentral chi-squared distribution with pq degrees of freedom and noncentrality parameter $\delta^2 ||M_1 + M_2^T||^2$, the asymptotic efficiencies are simply

$$\mathrm{ARE}(T_n, W_n) = \frac{||c_1 M_1 + c_2 M_2^T||^2}{pq\sigma_a^2\sigma_b^2||M_1 + M_2^T||^2},$$

where c_1 and c_2 are given in Theorem 3.3. Note that for multivariate normal distribution, $\psi(r) = r$, for k-variate t distribution with ν degrees of freedom, $\psi(r) = (k+\nu)r/(\nu + r^2)$ and for k-variate contaminated normal distribution with cdf $F(\boldsymbol{x}) = (1-\epsilon)\Phi(\boldsymbol{x}) + \epsilon\Phi(c^{-1}\boldsymbol{x})$, where $c > 0$ and Φ is the cdf of $N_k(\boldsymbol{0}, I_k)$,

$$\psi(r) = \frac{(1-\epsilon)\exp(-r^2/2) + \epsilon c^{-k-2}\exp(-r^2/2c^2)}{(1-\epsilon)\exp(-r^2/2) + \epsilon c^{-k}\exp(-r^2/2c^2)}\, r.$$

Assume now for simplicity that $M_1 = M_2^T$. For the limiting efficiency of the sign test of independence, we refer to Taskinen et al. (2003). The limiting efficiency of the Wilcoxon test T_{1n} with respect to the Wilks' test W_n is

$$\mathrm{ARE}(T_{1n}, W_n) = \frac{9(c_1 + c_2)^2}{4pq},$$

where

$$c_1 = E[G_1(r_i^{(1)})\psi_1(r_i^{(1)})]E[G_2(r_i^{(2)})r_i^{(2)}]$$

and

$$c_2 = E[G_2(r_i^{(2)})\psi_2(r_i^{(2)})]E[G_1(r_i^{(1)})r_i^{(1)}].$$

The resulting efficiencies for t distributions with selected degrees of freedom and dimensions are listed in Table 1 and for contaminated normal distributions with $\epsilon = 0.1$ and for selected values of c in Table 2. The efficiencies were derived using numerical integration.

Further, the limiting efficiency of the van der Waerden test T_{2n} as compared to the W_n is

$$\mathrm{ARE}(T_{2n}, W_n) = \frac{(c_1 + c_2)^2}{4p^2q^2},$$

TABLE 1. $ARE(T_{1n}, W_n)$ at different p- and q-variate t distributions for selected $\nu = \nu_1 = \nu_2$.

				p		
	q	2	3	5	8	10
	2	1.089	1.064	1.023	0.986	0.970
	3		1.039	0.998	0.961	0.946
$\nu = 5$	5			0.958	0.922	0.907
	8				0.886	0.871
	10					0.857
	2	0.970	0.960	0.934	0.907	0.893
	3		0.950	0.925	0.898	0.884
$\nu = \infty$	5			0.901	0.874	0.861
	8				0.848	0.835
	10					0.823

TABLE 2. $ARE(T_{1n}, W_n)$ at different p- and q-variate contaminated normal distributions for $\epsilon = 0.1$ and for selected values of c.

				p		
	q	2	3	5	8	10
	2	1.216	1.204	1.172	1.137	1.121
	3		1.192	1.161	1.126	1.109
$c = 3$	5			1.130	1.096	1.080
	8				1.063	1.048
	10					1.034
	2	1.833	1.815	1.767	1.714	1.689
	3		1.797	1.749	1.697	1.672
$c = 6$	5			1.703	1.652	1.628
	8				1.603	1.579
	10					1.556

where

$$c_1 = E\big\{[\Psi_p^{-1}(G_1(r_i^{(1)}))]^{1/2}\psi_1(r_i^{(1)})\big\}E\big\{[\Psi_q^{-1}(G_2(r_i^{(2)}))]^{1/2}r_i^{(2)}\big\}$$

and

$$c_2 = E\big\{[\Psi_q^{-1}(G_2(r_i^{(2)}))]^{1/2}\psi_2(r_i^{(2)})\big\}E\big\{[\Psi_p^{-1}(G_1(r_i^{(1)}))]^{1/2}r_i^{(1)}\big\}.$$

Now for the multivariate normal distribution, $ARE(T_{2n}, W_n) = 1$, and for the contaminated normal distribution, $ARE(T_{2n}, W_n) = (1 - \epsilon + \epsilon/c)^2(1 - \epsilon + \epsilon c)^2$. These efficiencies do not depend on the dimensions at all. For the efficiencies at certain contaminated normal distributions, see Figure 1. The efficiencies for t

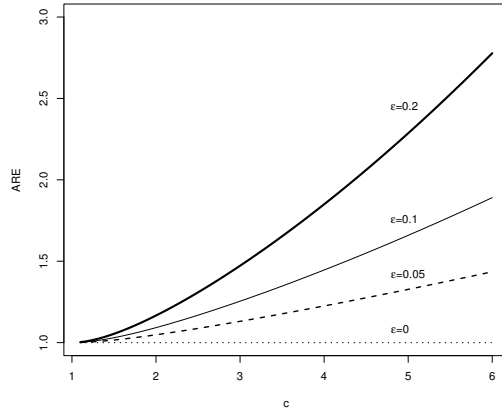distribution with 5 degrees of freedom were derived using numerical integration and are listed in Table 3.

FIGURE 1. $\mathrm{ARE}(T_{2n}, W_n)$ as a function of c at the contaminated normal model with $\epsilon = 0, 0.05, 0.10, 0.20$.

TABLE 3. $ARE(T_{2n}, W_n)$ at different p- and q-variate t distributions with $\nu_1 = \nu_2 = 5$.

q	p				
	2	3	5	8	10
2	1.125	1.132	1.144	1.155	1.160
3		1.140	1.151	1.162	1.168
5			1.162	1.174	1.179
8				1.185	1.190
10					1.195

Now some comments follow. First of all, the limiting efficiencies of the Wilcoxon test T_{1n} decrease with increasing dimension while the efficiencies of sign test T_{0n} and van der Waerden test T_{2n} increase or stay constant. Due to this property, for low dimensions, the efficiencies of T_{1n} are higher than those of T_{0n}, but for high dimensions, T_{0n} outperforms T_{1n}. The van der Waerden scores test is the most efficient one in all considered cases. When the underlying distribution is multivariate normal, it is as efficient as the Wilks' test. When the distribution becomes heavy-tailed, the efficiencies are higher than those of T_{0n} and T_{1n} (for the contaminated normal distribution with $\epsilon = 0.1$ and $c = 3$ and $c = 6$, the

efficiencies of T_{2n} are 1.254 and 1.891). For comparisons of limiting efficiencies, see also Figures 2 and 3.

4.2. A simulation study

A simple simulation study was used to compare the finite sample efficiencies of W_n, T_{0n}, T_{1n} and T_{2n}. 1500 independent $\boldsymbol{x}^{(1)}$- and $\boldsymbol{x}^{(2)}$-samples of sizes $n = 50$ and 200 were generated from a multivariate standard normal distribution, from a t distribution with 5 degrees of freedom and from a contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$. The transformation in (3.1) with $M_1 = M_2^T = I$ was applied for chosen values of $\Delta = \delta/\sqrt{n}$ to introduce dependence into the model. The tests were applied using the location and shape estimates chosen to satisfy

$$\text{ave}\{\widehat{\boldsymbol{S}}_i^{(1)}\} = \boldsymbol{0} \ \text{ and } \ p\,\text{ave}\{\widehat{\boldsymbol{S}}_i^{(1)}\widehat{\boldsymbol{S}}_i^{(1)^T}\} = I_p$$

and

$$\text{ave}\{\widehat{\boldsymbol{S}}_i^{(2)}\} = \boldsymbol{0} \ \text{ and } \ q\,\text{ave}\{\widehat{\boldsymbol{S}}_i^{(2)}\widehat{\boldsymbol{S}}_i^{(2)^T}\} = I_q,$$

that is, the transformation retransformation spatial median and the Tyler's M-estimate (Tyler, 1987; Hettmansperger and Randles, 2002). For the transformation retransformation technique, see also Chakraborty et al. (1998). The critical values used in test constructions were based on the chi-square approximations to the null distributions.

In Figure 2, the empirical powers as well as exact limiting powers ($n = \infty$) computed using Theorem 3.3 are given for $p = q = 3$. In the multivariate normal case W_n is slightly better than T_{1n} and T_{2n} and much better than T_{0n}. In the t distribution case no big differences can be seen between tests and in the contaminated normal case T_{1n} and T_{2n} outperform W_n and T_{0n}. In Figure 3, the empirical powers are illustrated for $p = q = 8$. In the multivariate normal case T_{0n} and T_{2n} are slightly more powerful than T_{1n}. In the considered t distribution case T_{1n} performs poorly, but as the underlying distribution is contaminated normal, T_{1n} performs very well. As $p = q = 8$, the sizes of T_{0n} and T_{2n} are often slightly below 0.05. The size of T_{1n} is very close to 0.05 in all cases and for heavy-tailed distributions, the size of W_n often exceeds 0.05.

5. A robustness study and final comments

Finally, a simple simulation study was used to illustrate the robustness of test statistics proposed above. Independent $\boldsymbol{x}^{(1)}$- and $\boldsymbol{x}^{(2)}$-samples of size $n = 30$ were generated from a bivariate standard normal distribution and the transformation in (3.1) with $M_1 = M_2^T = I_2$ was applied for chosen values of Δ to introduce "positive" dependence into the model. By positive dependence we mean that each $\boldsymbol{x}^{(1)}$-coordinate is positively dependent on each $\boldsymbol{x}^{(2)}$-coordinate. Finally, the first observation vectors in each sample were replaced by contaminated vectors $\boldsymbol{x}_1^{(1)} = (c, c)^T$ and $\boldsymbol{x}_1^{(2)} = (-c, -c)^T$ with "negative" dependence. The procedure was repeated 1000 times and mean p-values were computed.
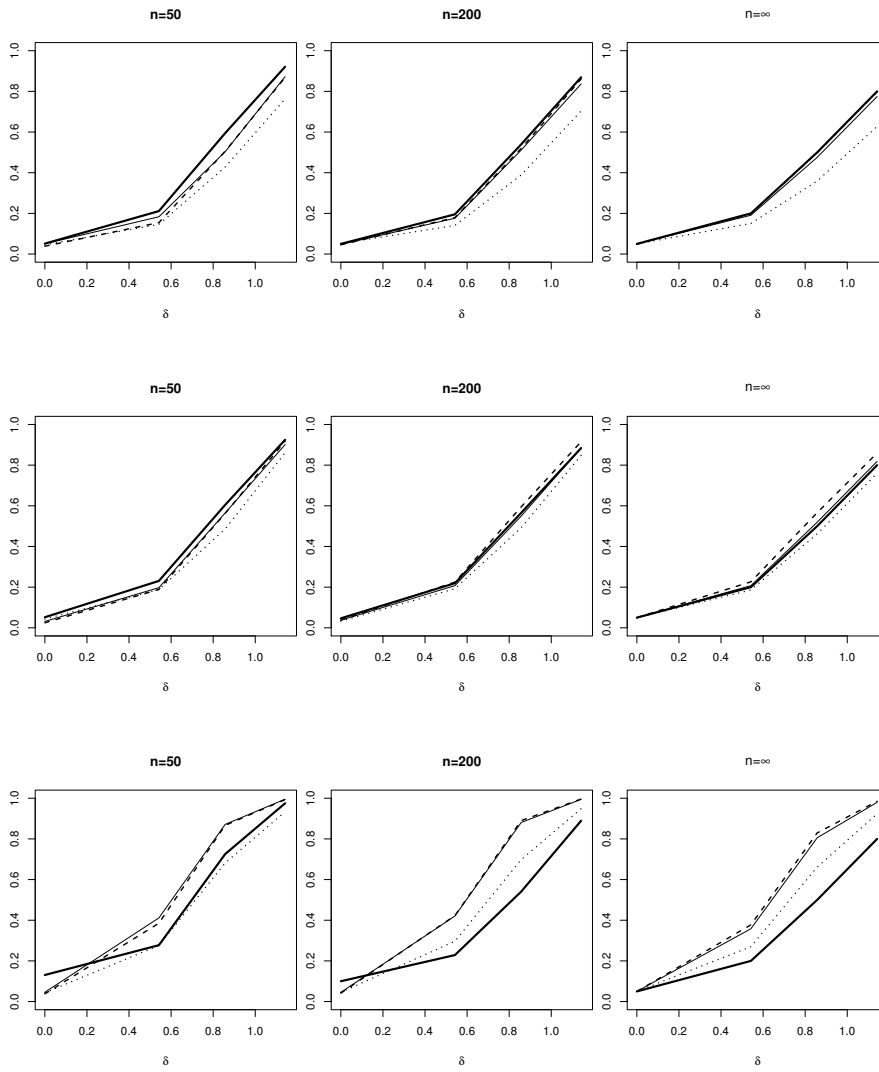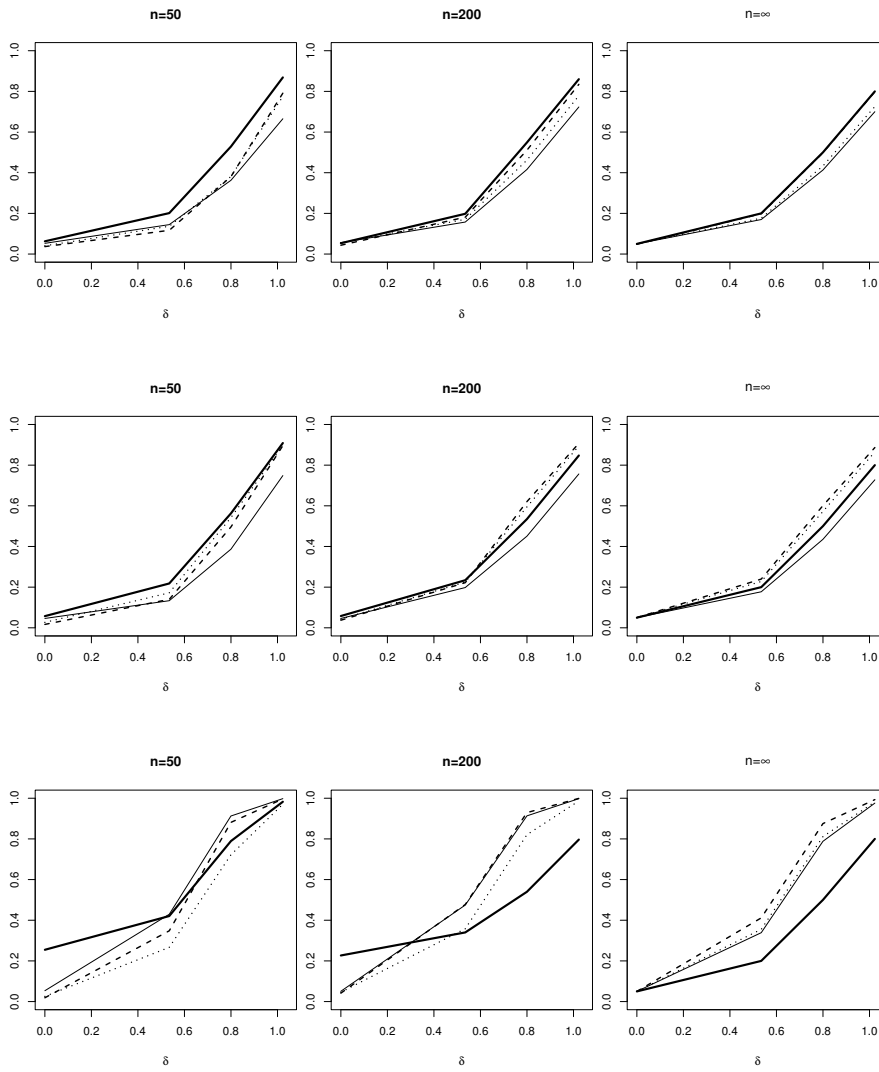
FIGURE 2. Empirical powers for $p = q = 3$ using the multivariate normal distribution (first row), multivariate t distribution with $\nu = 5$ (second row) and contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$ (third row). The thick solid line denotes W_n, the thin solid line T_{1n}, the thick dotted line T_{2n} and the thin dotted line T_{0n}.

FIGURE 3. Empirical powers for $p = q = 8$ using the multivariate
normal distribution (first row), multivariate t distribution with
$\nu = 5$ (second row) and contaminated normal distribution with
$\epsilon = 0.1$ and $c = 6$ (third row). The thick solid line denotes W_n, the
thin solid line T_{1n}, the thick dotted line T_{2n} and the thin dotted
line T_{0n}.

In Figure 4, the mean p-values are illustrated as a function of contamination value c for $\Delta = 0$ and for $\Delta = 0.2$. In the null hypothesis case ($\Delta = 0$), all tests give p-values close to 0.5, as the contamination value is near zero. Note also that T_{0n} and T_{1n} give practically the same p-values as Wilks' test. When the contamination value is high, p-values given by Wilks' test decrease considerably and some decrease is also seen in the p-values of T_{1n} and T_{2n}. In the considered case, the sign test T_{0n} seems to be the most robust one, since the mean p-value is constant as a function of c. As $\Delta = 0.2$, the contamination slightly increases the mean values of rank scores tests. In the case of Wilks' test the p-values first increase and then decrease to zero with the contamination value. The careful analysis shows that the small p-values for large contamination values erroneously indicate "negative" dependence, however.
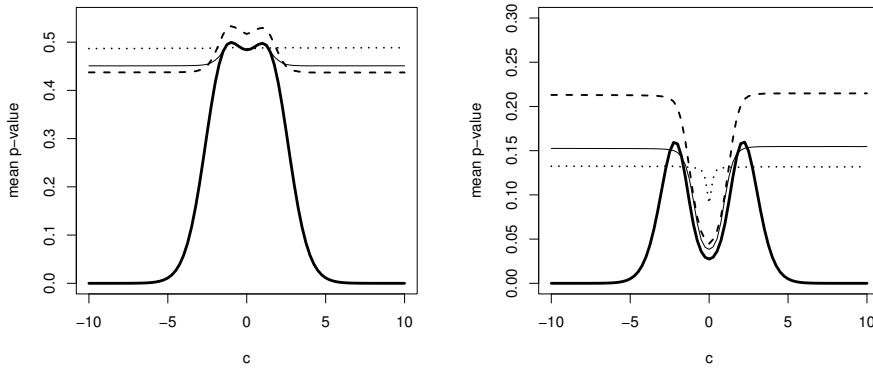

FIGURE 4. Mean p-values for the true null hypothesis $H_0 : \Delta = 0$ (left figure) and for the alternative hypothesis $H_1 : \Delta = 0.2$ (right figure) as a function of contamination value as described in the text. The thick solid line refers to W_n, the thin solid line to T_{1n}, the thick dotted line to T_{2n} and the thin dotted line to T_{0n}.

In the paper, new affine invariant rank scores procedures were proposed for testing whether two random vectors are independent. The test statistics were constructed using standardized spatial signs and ranks of the lengths of the standardized vectors. It is remarkable that, the proposed tests are valid without any moment assumptions on the underlying data as far as the standardization is done using such location vector and scatter (or shape) matrix estimates that do not require any moment assumptions. In the paper, three different score functions, namely the sign scores, the Wilcoxon scores and the van der Waerden scores were considered in more detail. The tests have good limiting and finite-sample efficiencies and as illustrated by an example, the tests are resistant to outliers.

Acknowledgements

Appendix I: Proofs of the results

Proof of Lemma 3.1   As T_n is affine invariant, we can without loss of generality restrict to the spherical case with $\boldsymbol{\mu} = \mathbf{0}$ and $C = I$. Assume now that the location and scatter estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{C}$ are $\sqrt{n}$-consistent and write $\boldsymbol{\mu}^* = \sqrt{n}\widehat{\boldsymbol{\mu}}$ and $C^* = \sqrt{n}(\widehat{C} - I)$.

The proof proceeds step by step as follows.

1. Assume that $||\boldsymbol{\mu}^*|| + ||C^*|| \leq \Delta$.
2. Then

$$\widehat{r}_i = r_i + \frac{1}{\sqrt{n}}s_i,$$

   where $|s_i| \leq M_s$, if $r_i \leq M_r$, and

$$\widehat{\boldsymbol{u}}_i = \boldsymbol{u}_i + \frac{1}{\sqrt{n}}\boldsymbol{v}_i,$$

   where $||\boldsymbol{v}_i|| \leq M_v$, if $r_i \geq m_r$.
3. Write $\widehat{G}_n(r) = (n+1)^{-1}\sum_j I(\widehat{r}_j \leq r)$ for the estimated cdf of the standardized radius. Then $\widehat{G}_n(\widehat{r}_i) = \widehat{R}_i/(n+1)$ and $\widehat{G}_n(r) \to_p G(r)$ uniformly in r. Moreover,

$$\widehat{G}_n(\widehat{r}_i) - G(r_i) \to_p 0 \ \text{ and also } \ \widehat{G}_n(\widehat{r}_i) - G(r_i) \xrightarrow{L_2} 0.$$

4. As $a(u)$ is continuous, monotone and square integrable, also

$$a(\widehat{G}_n(\widehat{r}_i)) - a(G(r_i)) \xrightarrow{L_2} 0.$$

   This is seen as

$$[a(\widehat{G}_n(\widehat{r}_i)) - a(G(r_i))]I(r_i \leq M_r) \xrightarrow{L_2} 0 \ \ \forall M_r$$

   and by Minkowski's inequality

$$\left\{E\left[[a(\widehat{G}_n(\widehat{r}_i)) - a(G(r_i))]I(r_i > M_r)\right]^2\right\}^{1/2}$$
$$\leq \left\{E\left[a(\widehat{G}_n(\widehat{r}_i))I(r_i > M_r)\right]^2\right\}^{1/2} + \left\{E\left[a(G(r_i))I(r_i > M_r)\right]^2\right\}^{1/2}$$
$$\to 2\left[\int_{G(M_r)}^1 a^2(u)du\right]^{1/2}$$

can be made as small as one wishes. For the latter convergence note that

$$E[a^2(\widehat{G}_n(\widehat{r}_i))I(r_i > M_r)]$$

$$\leq \frac{1}{n}\sum_i E[a^2(\widehat{G}_n(\widehat{r}_i))I(\widehat{G}_n(\widehat{r}_i) \geq \widehat{G}_n(M_r - M_s/\sqrt{n}))]$$

$$\leq E\left[\int_{\widehat{G}_n(M_r-M_s/\sqrt{n})}^{1} a^2(u)du\right] \rightarrow \int_{G(M_r)}^{1} a^2(u)du.$$

5. Next decompose $\widehat{H} - H$ into two parts as follows.

$$\sqrt{n}(\widehat{H} - H)$$

$$= \sqrt{n}\,\mathrm{ave}_i\left\{\left(a\left(\frac{\widehat{R}_i^{(1)}}{n+1}\right)b\left(\frac{\widehat{R}_i^{(2)}}{n+1}\right) - a(G_1(r_i^{(1)}))b(G_2(r_i^{(2)}))\right)\widehat{\boldsymbol{u}}_i^{(1)}\widehat{\boldsymbol{u}}_i^{(2)^T}\right\}$$

$$+ \sqrt{n}\,\mathrm{ave}_i\{a(G_1(r_i^{(1)}))b(G_2(r_i^{(2)}))(\widehat{\boldsymbol{u}}_i^{(1)}\widehat{\boldsymbol{u}}_i^{(2)^T} - \boldsymbol{u}_i^{(1)}\boldsymbol{u}_i^{(2)^T})\} =: H_1 + H_2.$$

So it is enough to show that $H_1 \rightarrow_p 0$ and $H_2 \rightarrow_p 0$. We proceed by proving $E[\mathrm{vec}(H_i)] \rightarrow 0$ and $Var[\mathrm{vec}(H_i)] \rightarrow 0$ for $i = 1, 2$.

6. As the standardized sign vectors are equivariant and ranks (of distances) are invariant under sign changes of the original data vectors, $E[\mathrm{vec}(H_1)] = 0$ and $E[\mathrm{vec}(H_2)] = 0$.

7. As

$$E\left[a\left(\frac{\widehat{R}_i^{(1)}}{n+1}\right) - a(G_1(r_i^{(1)}))\right]^2 \rightarrow 0, \quad E\left[b\left(\frac{\widehat{R}_i^{(2)}}{n+1}\right) - b(G_2(r_i^{(2)}))\right]^2 \rightarrow 0$$

and

$$\widehat{\boldsymbol{u}}_i^{(1)}\widehat{\boldsymbol{u}}_i^{(2)^T} = \boldsymbol{u}_i^{(1)}\boldsymbol{u}_i^{(2)^T} + \frac{1}{\sqrt{n}}\boldsymbol{v}_i^{(1)}\boldsymbol{u}_i^{(2)^T} + \frac{1}{\sqrt{n}}\boldsymbol{u}_i^{(1)}\boldsymbol{v}_i^{(2)^T} + o(1/\sqrt{n}),$$

it follows that, when $E[1/r_i^{(1)}] < \infty$ and $E[1/r_i^{(2)}] < \infty$,

$$E[\mathrm{vec}(H_1)\mathrm{vec}^T(H_1)]$$

$$= \frac{1}{n}\sum_i\sum_j E\left\{\left[a\left(\frac{\widehat{R}_i^{(1)}}{n+1}\right)b\left(\frac{\widehat{R}_i^{(2)}}{n+1}\right) - a(G_1(r_i^{(1)}))b(G_2(r_i^{(2)}))\right]\right.$$

$$\cdot\left[a\left(\frac{\widehat{R}_j^{(1)}}{n+1}\right)b\left(\frac{\widehat{R}_j^{(2)}}{n+1}\right) - a(G_1(r_j^{(1)}))b(G_2(r_j^{(2)}))\right]$$

$$\left.\cdot[\mathrm{vec}(\widehat{\boldsymbol{u}}_i^{(1)}\widehat{\boldsymbol{u}}_i^{(2)^T})\mathrm{vec}^T(\widehat{\boldsymbol{u}}_j^{(1)}\widehat{\boldsymbol{u}}_j^{(2)^T})]\right\} \rightarrow 0.$$

8. Consider next the variance of $\text{vec}(H_2)$. To shorten notations, write $a_i = a(G_1(r_i^{(1)}))$ and $b_i = b(G_2(r_i^{(2)}))$. The variance can then be written as

$$E[\text{vec}(H_2)\text{vec}^T(H_2)]$$

$$= \frac{1}{n^2}\sum_i\sum_j E[a_i b_i a_j b_j \text{vec}(\boldsymbol{u}_i^{(1)}\boldsymbol{v}_i^{(2)^T} + \boldsymbol{v}_i^{(1)}\boldsymbol{u}_i^{(2)^T})$$

$$\cdot \text{vec}^T(\boldsymbol{u}_j^{(1)}\boldsymbol{v}_j^{(2)^T} + \boldsymbol{v}_j^{(1)}\boldsymbol{u}_j^{(2)^T})] + o(1/\sqrt{n})$$

$$= \frac{1}{n^2}\sum_i E[a_i^2 b_i^2 \text{vec}(\boldsymbol{u}_i^{(1)}\boldsymbol{v}_i^{(2)^T} + \boldsymbol{v}_i^{(1)}\boldsymbol{u}_i^{(2)^T})$$

$$\cdot \text{vec}^T(\boldsymbol{u}_i^{(1)}\boldsymbol{v}_i^{(2)^T} + \boldsymbol{v}_i^{(1)}\boldsymbol{u}_i^{(2)^T})] + o(1/\sqrt{n}),$$

which converges (use again the sign change property) to zero when $E[1/r_i^{(1)}] < \infty$ and $E[1/r_i^{(2)}] < \infty$.

9. The result follows as $\boldsymbol{\mu}^*$ and C^* are bounded in probability.

Proof of Theorem 3.2   By Lemma 3.1, the limiting distribution of T_n can be found using the limiting distribution of $\sqrt{n}\text{vec}(H)$. Since for $i = 1, \ldots, n$,

$$E[a(G_1(r_i^{(1)}))b(G_2(r_i^{(2)}))\text{vec}(\boldsymbol{u}_i^{(1)}\boldsymbol{u}_i^{(2)^T})] = \boldsymbol{0}$$

and

$$E[a^2(G_1(r_i^{(1)}))b^2(G_2(r_i^{(2)}))\text{vec}(\boldsymbol{u}_i^{(1)}\boldsymbol{u}_i^{(2)^T})\text{vec}^T(\boldsymbol{u}_i^{(1)}\boldsymbol{u}_i^{(2)^T})] = \frac{\sigma_a^2\sigma_b^2}{pq}I_{pq},$$

where $\sigma_a^2 = E[a^2(U)]$ and $\sigma_b^2 = E[b^2(U)]$ with U uniformly distributed on $(0,1)$, the central limit theorem implies that $\sqrt{n}\text{vec}(H) \to_d N_{pq}(0, \sigma_a^2\sigma_b^2/pq\, I_{pq})$. Consequently,

$$T_n = \frac{npq}{\sigma_a^2\sigma_b^2}\|H\|^2 = \frac{npq}{\sigma_a^2\sigma_b^2}\text{vec}(H)^T\text{vec}(H) \xrightarrow{d} \chi_{pq}^2.$$

Proof of Theorem 3.3   In the proof, we apply LeCam's third lemma. See for example Hájek et al. (1999, Section 7.1). For testing H_0 against H_Δ, the optimal likelihood ratio test statistic is

$$L = \sum_{i=1}^{n}\{\log f_\Delta(\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)}) - \log f_0(\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)})\}.$$

Gieser (1993) considered the asymptotic representation $L = \sqrt{n}\delta K - \frac{1}{2}\delta^2\sigma^2 + o_P(1)$, where

$$K = \frac{1}{n}\sum_i k_i = \frac{1}{n}\sum_i\left[p - \psi_1(r_i^{(1)})r_i^{(1)} + \psi_1(r_i^{(1)})r_i^{(2)}\boldsymbol{u}_i^{(1)^T}M_1\boldsymbol{u}_i^{(2)}\right.$$

$$\left. + q - \psi_2(r_i^{(2)})r_i^{(2)} + \psi_2(r_i^{(2)})r_i^{(1)}\boldsymbol{u}_i^{(2)^T}M_2\boldsymbol{u}_i^{(1)}\right],$$

If in the above representation $E(k_i) = 0$ and $E(k_i^2) = \sigma^2$, the sequence of alternatives is contiguous to the null hypothesis (LeCam's first lemma). See Gieser (1993) for mild conditions.

Write then $\mathrm{vec}(H) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{h}_i$, where H is given in Lemma 3.1. We assume that, under H_0,

$$\sqrt{n}\begin{pmatrix} \mathrm{vec}(H) \\ K \end{pmatrix} \xrightarrow{d} N_{pq+1}\left(\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2\sigma_b^2/pq\, I_{pq} & E_0(\boldsymbol{h}_i k_i) \\ E_0^T(\boldsymbol{h}_i k_i) & E_0(k_i^2) \end{pmatrix} \right),$$

where E_0 denotes the expectations taken under the null hypothesis. Then by LeCam's third lemma, $\sqrt{n}\mathrm{vec}(H) \to_d N_{pq}(E_0(\boldsymbol{h}_i k_i), \sigma_a^2\sigma_b^2/pq\, I_{pq})$ under the alternative sequences.

Using the independence of $r_i^{(1)}$, $\boldsymbol{u}_i^{(1)}$, $r_i^{(2)}$ and $\boldsymbol{u}_i^{(2)}$, it is easy to see that

$$E_0(\boldsymbol{h}_i k_i) = \frac{\delta}{pq}\Bigg\{ E_0[a(G_1(r_i^{(1)}))\psi_1(r_i^{(1)})]E_0[b(G_2(r_i^{(2)}))r_i^{(2)}]\mathrm{vec}(M_1)$$

$$+ E_0[b(G_2(r_i^{(2)}))\psi_2(r_i^{(2)})]E_0[a(G_1(r_i^{(1)}))r_i^{(1)}]\mathrm{vec}(M_2^T)\Bigg\}$$

$$=: \frac{\delta}{pq}\Bigg\{ \mathrm{vec}(c_1 M_1 + c_2 M_2^T) \Bigg\}.$$

Hence under the alternative sequences,

$$\sqrt{n}\mathrm{vec}(H) \xrightarrow{d} N_{pq}\left(\frac{\delta}{pq}\mathrm{vec}(c_1 M_1 + c_2 M_2^T), \frac{\sigma_a^2\sigma_b^2}{pq} I_{pq} \right)$$

and

$$T_n^* = \frac{npq}{\sigma_a^2\sigma_b^2} \|H\|^2 \xrightarrow{d} \chi_{pq}^2\left(\frac{\delta^2}{pq\sigma_a^2\sigma_b^2}\|c_1 M_1 + c_2 M_2^T\|^2 \right).$$

Appendix II: Some notions on alternative sequences

For all elliptic cases, it is enough to consider the alternative sequences

$$\begin{pmatrix} \boldsymbol{y}_i^{(1)} \\ \boldsymbol{y}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (1-\Delta)I_p & \Delta M_1 \\ \Delta M_2 & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix},$$

where $\Delta = \delta/\sqrt{n}$ and $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ are independent with **spherical marginal distributions**. This is because, for the weighted sum of elliptical marginals,

$$\begin{pmatrix} \boldsymbol{y}_i^{(1)} \\ \boldsymbol{y}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (1-\Delta)I_p & \Delta M_1 \\ \Delta M_2 & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} A\boldsymbol{x}_i^{(1)} \\ B\boldsymbol{x}_i^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} (1-\Delta)I_p & \Delta A^{-1}M_1 B \\ \Delta B^{-1}M_2 A & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} (1-\Delta)I_p & \Delta M_1' \\ \Delta M_2' & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix}.$$

Hence (due to affine invariance) one can as well consider the sequence

$$\begin{pmatrix} (1-\Delta)I_p & \Delta M_1' \\ \Delta M_2' & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix},$$

for different choices of M_1' and M_2'. In all cases, the efficiencies are then of the same type $||c_1 M_1 + c_2 M_2^T||$, where c_1 and c_2 depend on the marginal spherical distributions and the test used. If the marginal distributions are of the same type (that is $c_1 = c_2$ for all tests to be compared), then the efficiencies do not depend on M_1 and M_2. Note also that the tests are not "unbiased" (noncentrality parameter > 0) for all alternative sequences. They are all unbiased for normal marginals. If the marginals are nonnormal but of the same type, then they all fail under $M_1 = -M_2^T$.

References

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

[2] B. Chakraborty, P. Chaudhuri, and H. Oja, *Operating transformation retransformation on spatial median and angle test.* Statistica Sinica **8** (1998), 767–784.

[3] P.W. Gieser, *A New Nonparametric Test of Independence Between Two Sets of Variates.* (1993), Unpublished Ph.D. dissertation, University of Florida, Gainesville. (http://www.cyberpete.com/thesis.do)

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

[5] J. Hájek, Z. Šidák and P.K. Sen, *Theory of Rank Tests.* Academic Press, (1999).

[6] M. Hallin and D. Paindaveine, *Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks.* Annals of Statistics **30** (2002), 1103–1133.

[7] T.P. Hettmansperger and R.H. Randles, *A practical affine equivariant multivariate median.* Biometrika **89** (2002), 851–860.

[8] E. Ollila, T.P. Hettmansperegr and H. Oja, *Affine equivariant multivariate sign methods.* (2003), under revision.

[9] M.L. Puri and P.K. Sen, *Nonparametric Methods in Multivariate Analysis.* Wiley, New York, (1971).

[10] S. Taskinen, A. Kankainen and H. Oja, *Sign test of independence between two random vectors.* Stat. Prob. Letters **62** (2003), 9–21.

[11] S. Taskinen, H. Oja and R.H. Randles, *Multivariate nonparametric tests of independence.* (2005), To appear in *JASA*.

[12] D.E. Tyler, *A distribution-free M-estimator of multivariate scatter.* Ann. Statist. **15** (1987), 234–251.

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

Department of Mathematics and Statistics, University of Jyväskylä, P.O.Box 35, FIN-40014 Jyväskylä, Finland
E-mail address: {slahola, kankaine, Hannu.Oja}@maths.jyu.fi