Austrian Journal of Statistics
July 2019, Volume 48, 14–42.
http://www.ajs.or.at/
doi:10.17713/ajs.v48i4.654

Nonparametric Tests Applicable to High Dimensional Data

František Rublík
Institute of Measurement Science, Slovak Academy of Sciences

Abstract

Constructions of data driven ordering of set of multivariate observations are presented. The methods employ also dissimilarity measures. The ranks are used in the construction of test statistics for location problem and in the construction of the corresponding multiple comparisons rule. An important aspect of the resulting procedures is that they can be used also in the multisample setting and in situations where the sample size is smaller than the dimension of the observations. The performance of the proposed procedures is illustrated by simulations.

Keywords: ranks of multivariate observations, tests of multivariate location hypotheses, high dimensional data, multiple comparisons for location parameters, projection pursuit.

1. Introduction

The nonparametric test statistics (based on ranks) for testing the equality of the location parameters of one-dimensional statistical populations have universal applicability. This is because in the continuous case under the validity of the null hypothesis they have exact distribution not depending on the sampled distribution and therefore in such a case they yield chosen significance level. Nonparametric test statistics aimed at testing the equality of location parameters of the multivariate populations usually provide the chosen significance level only in the asymptotic sense, i.e., they attain the significance level only in the limit as the sample sizes tend to infinity. Although simulations based on sampling from particular populations yield some picture of the coincidence of the size of such a test with the chosen significance level, the disadvantage is that the user does not have a guaranteed information about this coincidence in the case when the distribution does not belong to the probability families included in the simulation study.

In this paper ranks of multivariate observations are constructed and they are used in construction of tests of the equality of location parameters of multivariate populations. These tests are based on the use of the classical rank statistics. Under validity of the null hypothesis, similarly as in the classical one-dimensional case, for sampling from continuous multivariate population the proposed ranks are uniformly distributed, the resulting test statistics possess an exact null distribution and yield tests providing the chosen significance level. Moreover, these tests can be used for testing the multivariate location hypothesis also when the dimensionality of the

data is greater than the sample size.

A general construction of the ranks of multivariate vectors is the topic of Subsection 2.1, ranks constructed by means of a random vector are mentioned in Subsection 2.2, the main result of the paper, the ranks constructed by means of a dissimilarity, are presented in Subsection 2.3. The multivariate 2-sample Wilcoxon tests based on the results from Section 2 are described in Subsection 3.1, their simulation power comparison with known nonparametric 2-sample location tests is in Subsection 3.2. Multisample tests of location in the small sample setting are briefly mentioned in Subsection 3.3. All simulation results of the paper are based on the use of MATLAB. The conclusions are in Section 4, the computation of the Friedman projection pursuit vector based on the the spherical coordinates is mentioned in Section 4.

2. Ranks of multivariate observations

The existence of the exact null distribution of the classical rank statistics in the one-dimensional case follows from the fact the ranks of the random sample from continuous univariate distribution are uniformly distributed over the corresponding set of permutations. The ranks of real valued observations are based on ordering according to their magnitude. Some possible principles of ordering multivariate data are discussed by Barnett (1976). As there is not known any natural ordering of the multivariate Euclidean space yielding useful statistical procedures, the successes of ranking in the one-dimensional case suggest that some data driven ordering of elements of random sample of size n, i.e, not the 2-vector based but an n-vector based relation, could be helpful. In this paper both the ranks computed by means of the scalar product with a random vector, and ranks based on a dissimilarity, are presented. These constructions, similarly as in the one-dimensional case, lead to statistics having an exact null distribution.

2.1. A general construction

The purpose of this Subsection is to present a basis for definition of ranks in the multivariate case.

Let $\mathbb{R}^d$ denote the Euclidean space of d-dimensional column vectors. Throughout the paper we assume that $d > 1$ and the integer $n \geq 5$. We shall use the notation

$$\mathbb{R}^{(d,n)} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_n);\ \mathbf{x}_i \in \mathbb{R}^d, i = 1, \ldots, n\},$$
$$\tilde{\mathbb{R}}^{(d,n)} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{(d,n)};\ \mathbf{x}_1, \ldots, \mathbf{x}_n \textit{ are mutually different vectors}\}.$$

In this paper we shall often use the fact that if $\mathbf{c} \in \mathbb{R}^m$ is a non-zero vector, then the set of the $\mathbf{z}$'s for which $\mathbf{c}'\mathbf{z} = 0$, has the multivariate Lebesgue measure zero. By means of this it is obvious that the multivariate Lebesgue measure $\mu_L^{(d,n)}$ of the set $\mathbb{R}^{(d,n)} - \tilde{\mathbb{R}}^{(d,n)}$ equals zero.

As usual, by $\mathcal{R}^{(n)}$ we understand the set of all permutations of the integers $1, \ldots, n$.

Theorem 2.1.1. *Suppose that*

$$\mathcal{D}_n \subset \tilde{\mathbb{R}}^{(d,n)},\ (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n \textit{ whenever } (\mathbf{x}_{r_1}, \ldots, \mathbf{x}_{r_n}) \in \mathcal{D}_n \textit{ for some } r \in \mathcal{R}^{(n)},$$
$$\mu_L^{(d,n)}(\tilde{\mathbb{R}}^{(d,n)} - \mathcal{D}_n) = 0, \tag{2.1.1}$$

where $\mu_L^{(d,n)}$ denotes the multivariate Lebesgue measure on $\mathbb{R}^{(d,n)}$. Let for each $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ belonging to $\mathcal{D}_n$ there exist a unique n-tuple

$$\mathbf{x}^{(.)} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}) \in \mathcal{D}_n, \quad \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}\} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \tag{2.1.2}$$

such that $\mathbf{x}^{(.)}$ is a measurable mapping of the argument $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n), (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n) \in \mathcal{D}_n, \quad \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n\} \implies \mathbf{x}^{(i)} = \tilde{\mathbf{x}}^{(i)},\ i = 1, \ldots, n. \tag{2.1.3}$$

For $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n$ *define the vector of ranks* $R = R(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (R_1, \ldots, R_n)$ *with values in* $\mathcal{R}^{(n)}$ *by the formula*

$$R_i = R_i(\mathbf{x}) = R_i(\mathbf{x}_i) = r_i \ \textit{if} \ \mathbf{x}_i = \mathbf{x}^{(r_i)}. \tag{2.1.4}$$

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be a sequence of d-dimensional random vectors which has a density with respect to the multivariate Lebesgue measure. Then the random matrix* $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \mathcal{D}_n$ *with probability* 1 *and consequently, the vector of ranks* $R(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *is uniquely defined almost surely. Moreover, if in addition to this assumption the sequence* $(\mathbf{X}_{r_1}, \ldots, \mathbf{X}_{r_n})$ *has the same distribution for every permutation* $r \in \mathcal{R}^{(n)}$*, then for the vector of ranks* $R = R(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *the equality* $P(R = r) = 1/n!$ *holds for each permutation* $r \in \mathcal{R}^{(n)}$*.*

Proof. The proof is analogous to the proof for classical one-dimensional case $d = 1$, handled by Hájek, Šidák, and Sen (1999). Since (2.1.1) holds and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ has a density with respect to the multivariate Lebesgue measure, $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \mathcal{D}_n$ almost surely and consequently, $R(\mathbf{X})$ is uniquely defined almost surely. Obviously, $R : \mathcal{D}_n \longrightarrow \mathcal{R}^{(n)}$ is a measurable mapping. Now let us assume that the distribution of $(\mathbf{X}_{r_1}, \ldots, \mathbf{X}_{r_n})$ is the same for every permutation $r \in \mathcal{R}^{(n)}$. Let a permutation $r \in \mathcal{R}^{(n)}$ be arbitrary, but fixed. Put $\mathbf{Y}_i = \mathbf{X}_{r_i^{-1}}$, $i = 1, \ldots, n$. Then $\mathbf{X}$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ are equally distributed. Let $eid = (1, 2, \ldots, n)$ denote the identical permutation from $\mathcal{R}^{(n)}$. Since $R(\mathbf{X}) = r$ if and only $R(\mathbf{Y}) = eid$, we have

$$P(R(\mathbf{X}) = r) = P(R(\mathbf{Y}) = eid) = P(R(\mathbf{X}) = eid),$$

and by means of this the rest of the proof can be simply carried out. □

2.2. Ranks constructed by means of a random vector

In this Subsection we deal with construction of ranks by means of the scalar products of observations with a vector which is a function of these observations. Thus in difference from Fraiman, Moreno, and Vallejo (2017), where the multiplying vectors are chosen randomly and yield randomized tests with which various statisticians may come to various conclusions with the same data, the tests presented in this Subsection are not randomized.

Theorem 2.2.1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be a sequence of d-dimensional random vectors which has a density with respect to the multivariate Lebesgue measure. Suppose that* $\mathbf{c} : \mathbb{R}^{(d,n)} \to \mathbb{R}^d$ *is a measurable mapping and there exists a set* $\mathcal{D}_n$ *fulfilling (2.1.1) such that the equality* $\mathbf{c}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbf{c}(\mathbf{x}_{r_1}, \ldots, \mathbf{x}_{r_n})$ *holds for each* $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n$ *and each permutation* $r \in \mathcal{R}^{(n)}$*. For* $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n$ *define*

$$\begin{gathered}\mathbf{x}_{i_1} \prec \mathbf{x}_{i_2} \ \textit{if either} \ \mathbf{c}(\mathbf{x})^\top \mathbf{x}_{i_1} < \mathbf{c}(\mathbf{x})^\top \mathbf{x}_{i_2} \ \textit{or} \\ \mathbf{c}(\mathbf{x})^\top \mathbf{x}_{i_1} = \mathbf{c}(\mathbf{x})^\top \mathbf{x}_{i_2} \ \textit{and} \ \mathbf{x}_{i_1}(\ell) < \mathbf{x}_{i_2}(\ell), \ \ell = min\{k; |\mathbf{x}_{i_1}(k) - \mathbf{x}_{i_2}(k)| > 0\}.\end{gathered} \tag{2.2.1}$$

Let

$$\mathbf{x}^{(1)} \prec \mathbf{x}^{(2)} \prec \ldots \prec \mathbf{x}^{(n)} \tag{2.2.2}$$

denote this ordering of $\mathbf{x}_1, \ldots, \mathbf{x}_n$*. Then the relations (2.1.2), (2.1.3) hold and if the sequence* $(\mathbf{X}_{r_1}, \ldots, \mathbf{X}_{r_n})$ *has the same distribution for every permutation* $r \in \mathcal{R}^{(n)}$*, the vector of ranks* $R = R(\mathbf{X}_1, \ldots, \mathbf{X}_n) = (R_1, \ldots, R_n)$ *computed by means of (2.1.4) has the properties described in Theorem 2.1.1.*

Proof. Since $\mu_L^{(d,n)}(\mathbb{R}^{(d,n)} - \mathcal{D}_n) = 0$ by (2.1.1), and the random matrix $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ possesses a density with respect to $\mu_L^{(d,n)}$, the equality $P((\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \mathcal{D}_n) = 1$ holds. Hence denoting $\mathbf{x}^{(1)} \prec \cdots \prec \mathbf{x}^{(n)}$ the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathcal{D}_n$ in the increasing order, one can easily carry out the proof by means of Theorem 2.1.1. □

2.3. Ranks based on a dissimilarity

While the ranks constructed in the previous Subsection are based on the scalar product of a random vector with the observed data, now we are going to propose a construction of the ranks by means of a dissimilarity.

In the next procedure the order statistics $\mathbf{x}^{(.)}$ are constructed in such a way that we compute (step by step) lower indexes $(\ell_1, \dots, \ell_m)$ and upper indexes $(h_{n-m}, \dots, h_1)$. The result of the computation $(i_1, \dots, i_n) = (\ell_1, \dots, \ell_m, h_{n-m}, \dots, h_1)$ is a permutation of $1, \dots, n$, the order statistics will be $\mathbf{x}^{(j)} = \mathbf{x}_{i_j}$ and the ranks are determined by (2.1.4). In doing this also the nearest neighbors classification algorithm will be used. Its decision rule is based on a critical constant $k = k(N)$, where N is the number of the elements of the training set (more on this in the text following the next construction). This construction will be formulated in a general form by means of a dissimilarity measure.

By the dissimilarity measure we understand a function $t : \mathbb{R}^d \to \langle 0, \infty)$ such that $t(\mathbf{x}) = t(-\mathbf{x})$ for all $\mathbf{x}$ and $t(\mathbf{x}) = 0$ for $\mathbf{x} = \mathbf{0}$. Throughout the text for $\mathbf{z} = (z_1, \dots, z_d)^\top \in \mathbb{R}^d$

$$\mathbf{z}(j) = \pi_j(\mathbf{z}) = z_j \tag{2.3.1}$$

denotes the j-th coordinate of the vector.

Let us assume that t is a fixed dissimilarity and

$$\mathbf{u} : \mathbb{R}^d \to \mathbb{R}^d \tag{2.3.2}$$

is a measurable mapping. For $\mathbf{x}_1, \dots, \mathbf{x}_n$ belonging to $\mathbb{R}^d$, $n \geq 5$ compute

$$\overline{\mathbf{u}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_i, \quad \mathbf{V} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{u}_i - \overline{\mathbf{u}})(\mathbf{u}_i - \overline{\mathbf{u}})^\top, \tag{2.3.3}$$

where $\mathbf{u}_i = \mathbf{u}(\mathbf{x}_i)$. Let $\tilde{\mathbf{c}} = \tilde{\mathbf{c}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be a non-zero characteristic vector corresponding to the largest characteristic root of $\mathbf{V}$. Put

$$\mathbf{c} = \mathbf{c}(\tilde{\mathbf{c}}) = \begin{cases} \tilde{\mathbf{c}} & if \ \sum_{j=1}^{d} \tilde{\mathbf{c}}(j) > 0, \\ \tilde{\mathbf{c}} & if \ \sum_{j=1}^{d} \tilde{\mathbf{c}}(j) = 0 \ and \ \tilde{\mathbf{c}}(\tilde{j}) > 0, \ \tilde{j} = min\{j;\ |\tilde{\mathbf{c}}(j)| > 0\}, \\ -\tilde{\mathbf{c}} & otherwise. \end{cases} \tag{2.3.4}$$

As the observations may be subjected to some rounding of the values, for $\mathbf{x}_{i_1}, \mathbf{x}_{i_2} \in \mathbb{R}^d$ consider the following inequalities, which serve as a basis for the breaking rule (2.3.13)–(2.3.16).

$$\sum_{j=1}^{d} \pi_j\Big(\mathbf{u}(\mathbf{x}_{i_1})\Big) \neq \sum_{j=1}^{d} \pi_j\Big(\mathbf{u}(\mathbf{x}_{i_2})\Big), \tag{2.3.5}$$

$$\sum_{j=1}^{d} \pi_j(\mathbf{x}_{i_1}) \neq \sum_{j=1}^{d} \pi_j(\mathbf{x}_{i_2}), \tag{2.3.6}$$

$$\|\mathbf{x}_{i_1}\| \neq \|\mathbf{x}_{i_2}\|, \tag{2.3.7}$$

$$\pi_j(\mathbf{x}_{i_1}) \neq \pi_j(\mathbf{x}_{i_2}) \ for \ some \ j. \tag{2.3.8}$$

Here $\|\mathbf{x}_i\| = \sqrt{\mathbf{x}_i'\mathbf{x}_{i_1}}$ denotes the usual Euclidean norm. In the construction of the ordering of the data we shall use the constraint

$$the \ numbers \ \mathbf{c}^\top \mathbf{u}_i, \ i = 1, 2, \dots, n \ are \ mutually \ different \tag{2.3.9}$$

and put

$$\mathcal{D}_n^{(sp)} = \Big\{(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{(d,n)};\ either \ (2.3.9) \ is \ fulfilled \ or \ for \ each \ integers \ 1 \leq i_1 < i_2 \leq n \\ some \ of \ the \ inequalities \ (2.3.5) \ - \ (2.3.8) \ holds\Big\} \tag{2.3.10}$$

where the superscript (sp) is used to express that the constraints (2.3.9) concern the special chosen mapping (2.3.2). Assume that $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n^{(sp)}$ and compute the matrix of dissimilarities

$$\mathbf{dM}(i,j) = \begin{cases} t(\mathbf{u}_i - \mathbf{u}_j) & i \neq j, \\ 0 & i = j. \end{cases} \tag{2.3.11}$$

Step 1. Put

$$\begin{gathered} \mathbf{c}^\top \mathbf{u}_i \prec \mathbf{c}^\top \mathbf{u}_j \ \textit{if} \ \ \mathbf{c}^\top \mathbf{u}_i < \mathbf{c}^\top \mathbf{u}_j \\ \textit{or } \mathbf{c}^\top \mathbf{u}_i = \mathbf{c}^\top \mathbf{u}_j \ \textit{and one of the following possibilities holds:} \end{gathered} \tag{2.3.12}$$

$$\sum_k \pi_k(\mathbf{u}_i) < \sum_k \pi_k(\mathbf{u}_j), \tag{2.3.13}$$

$$\sum_k \pi_k(\mathbf{u}_i) = \sum_k \pi_k(\mathbf{u}_j) \ \textit{and} \ \sum_k \pi_k(\mathbf{x}_i) < \sum_k \pi_k(\mathbf{x}_j), \tag{2.3.14}$$

$$\sum_k \pi_k(\mathbf{u}_i) = \sum_k \pi_k(\mathbf{u}_j), \ \sum_k \pi_k(\mathbf{x}_i) = \sum_k \pi_k(\mathbf{x}_j) \ \textit{and} \ \|\mathbf{x}_i\| < \|\mathbf{x}_j\|, \tag{2.3.15}$$

$$\left.\begin{gathered} \sum_k \pi_k(\mathbf{u}_i) = \sum_k \pi_k(\mathbf{u}_j), \ \sum_k \pi_k(\mathbf{x}_i) = \sum_k \pi_k(\mathbf{x}_j), \\ \|\mathbf{x}_i\| = \|\mathbf{x}_j\| \ \textit{and} \ \pi_\ell(\mathbf{x}_i) < \pi_\ell(\mathbf{x}_j), \ \ell = \min\{k; \pi_k(\mathbf{x}_i) \neq \pi_k(\mathbf{x}_j). \end{gathered}\right\} \tag{2.3.16}$$

It follows from the definition of the set $\mathcal{D}_n^{(sp)}$ that this binary relation orders these scalar products linearly. Thus we obtain the ordering

$$\mathbf{c}^\top \mathbf{u}_{i_1} \prec \mathbf{c}^\top \mathbf{u}_{i_2} \prec \ldots \prec \mathbf{c}^\top \mathbf{u}_{i_n} \tag{2.3.17}$$

of the numbers $\mathbf{c}^\top \mathbf{u}_i$. By means of this compute

$$m = max(0.1n, 2), \quad low = (i_1, i_2, \ldots, i_m), \quad up = (i_{n-m+1}, i_{n-m+2}, \ldots, i_n). \tag{2.3.18}$$

Step II. Loosely speaking, the new set *low* will be constructed by adding index j of the observation $\mathbf{u}_j$ (classified by the nearest neighbors rule as belonging to the observations with indexes from *low*) which is the most distant from the upper set, to the indexes in *low*. The new upper set will be constructed analogically.

To do this, suppose that we have already constructed the vectors of lower and upper indexes

$$low = (i_1, \ldots, i_s), \quad up = (i_{n-h+1}, i_{n-h+2}, \ldots, i_n).$$

The corresponding lower and upper observations are

$$\mathbf{U}_{low} = (\mathbf{u}_{i_1}, \ldots, \mathbf{u}_{i_s}), \quad \mathbf{U}_{up} = (\mathbf{u}_{i_{n-h+1}}, \ldots, \mathbf{u}_{i_n}) \tag{2.3.19}$$

and the set of the remaining indexes

$$rem = \{j \in \{1, \ldots, n\}; j \notin (low \bigcup up)\} = \{j_1, \ldots, j_g\}, \quad g = n - (s + h).$$

Hence $s + h$ is the number of the chosen indexes, which will be used in the following classification in the training sets. For $i = 1, \ldots, g$ compute the result of the nearest neighbors classification rule

$$clas(j_i) = clasNN\Big(k(s+h), \mathbf{U}_{low}, \mathbf{U}_{up}, \mathbf{u}_{j_i}\Big),$$

based on the computed dissimilarities (2.3.11) (a detailed explanation is in the text following the end of this construction). Thus the index j_i is classified to the lower indexes if $clas(j_i) = 1$ and to the upper indexes if $clas(j_i) = 2$. Further, compute (cf. (2.3.11))

$$preLow = \{j \in rem; \, clas(j) = 1\}, \quad sumUp(j) = \sum_{i \in up} \mathbf{dM}(j, i)$$

and find the set

$$J_{low} = \{j^* \in preLow;\, sumUp(j^*) = \max_{j \in preLow} sumUp(j)\}.$$

Let $J_{low} = \{j_1^*, \ldots, j_\ell^*\}$. Write $JJ_{low} = (j_1^*)$ if $\ell = 1$, otherwise order the set J_{low} in such way that (cf. (2.3.17))

$$\mathbf{c}^\top \mathbf{u}_{j_1^*} \prec \ldots \prec \mathbf{c}^\top \mathbf{u}_{j_\ell^*}$$

and put $JJ_{low} = j_1^*$.

Similarly, let

$$preUp = \{j \in rem;\, clas(j) = 2\}, \quad sumLo(j) = \textstyle\sum_{i \in low} \mathbf{dM}(j,i),$$
$$J_{up} = \{\tilde{j} \in preUp;\, sumLo(\tilde{j}) = \max_{j \in preLow} sumLo(j)\},$$
$$J_{up} = \{\tilde{j}_1, \ldots, \tilde{j}_a\}, \quad \mathbf{c}^\top \mathbf{u}_{\tilde{j}_1} \prec \ldots \prec \mathbf{c}^\top \mathbf{u}_{\tilde{j}_a}$$
$$JJ_{up} = \tilde{j}_a,$$

of course if the set J_{up} consists of one element then $a = 1$.

Now define

$$newLow = \begin{cases} (low, JJ_{low}) & preLow \neq \emptyset, \\ low & otherwise, \end{cases}$$

$$newUp = \begin{cases} (JJ_{up}, up) & preUp \neq \emptyset, \\ up & otherwise. \end{cases}$$

To finish this step, put

$$low = newLow, \qquad up = newUp.$$

Step III. Repeat Step II while the number of the elements in (low, up) is less than n.

After finishing this procedure we obtain the vector $(low, up) = (i_1, \ldots, i_n)$, which is a permutation of $1, \ldots, n$. The resulting order statistics are

$$\mathbf{x}^{(.)} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}), \qquad \mathbf{x}^{(j)} = \mathbf{x}_{i_j},\ j = 1, \ldots, n. \tag{2.3.20}$$

To avoid possible ambiguities in the further use of this construction let us explain in detail the nearest neighbors classification rule used in the previous text. Let $low = (i_1, \ldots, i_s)$, $up = (i_{n-h+1}, i_{n-h+2}, \ldots, i_n)$, z be mutually different integers from the set $\{1, \ldots, n\}$ and k is a positive integer. The dissimilarity t based nearest neighbors classification (cf. (2.3.19))

$$y = clasNN\Big(k, \mathbf{U}_{low}, \mathbf{U}_{up}, \mathbf{u}_z\Big) \tag{2.3.21}$$

of the vector $\mathbf{u}_z$ is defined as follows. Use the matrix (2.3.11) and put $v_i = \mathbf{dM}(i,z)$, $i \in low \bigcup up$. Let $\mathbf{c}$ be the vector from (2.3.4). Define

$$v_i \prec v_j \ \textit{if}\ \ v_i < v_j \ \textit{or}\ v_i = v_j \ \textit{and}\ \mathbf{c}^\top \mathbf{u}_i \prec \mathbf{c}^\top \mathbf{u}_j \ \textit{in the sense (2.3.12) – (2.3.16).} \tag{2.3.22}$$

It follows from the definition of the set $\mathcal{D}_n^{(sp)}$ that this binary relation orders the numbers v_i, $i \in low \bigcup up$, linearly. Thus we obtain the ordering

$$v^{(1)} \prec v^{(2)} \prec \ldots \prec v^{(s+h)}$$

of the numbers v_i. Use the constant k postulated in (2.3.21) and put

$$h(1) = \#\{f \in \{1, \ldots, k\};\, v^{(f)} = v_i \textit{ for some } i \in low\,\},$$
$$h(2) = \#\{f \in \{1, \ldots, k\};\, v^{(f)} = v_i \textit{ for some } i \in up\,\}.$$

Then $y = 1$ if $h(1) > h(2)$ (i.e, in this case $\mathbf{u}_z$ is added to $\mathbf{U}_{low}$), and $y = 2$ if $h(1) < h(2)$ (now $\mathbf{u}_z$ is added to $\mathbf{U}_{up}$). If $h(1) = h(2)$, use this procedure with $k + 1$ instead of k.

The constant k is in the rule (2.3.21) chosen in dependence on the number $N = s + h$ of the elements $\mathbf{u}_i$, $i \in low \bigcup up$, of the training set. In general, there are no tables containing a prescription how to choose this constant. It is known that it should be chosen in such a way that $k(N)/N \to 0$ as $N \to \infty$. However, if this rule has to be implemented in a situation with particular data, one needs a prescription determining the choice of k. In the simulations mentioned in the next sections there are used the constants $k(N)$ defined as follows. Obviously,

Table 2.3.1: Values of critical constant k used in the nearest neighbors rule for $2 \le N \le 400$.

N	$2 \le N \le 6$	$7 \le N \le 11$	$12 \le N \le 17$	$18 \le N \le 31$	$32 \le N \le 41$
k(N)	2	3	4	5	6
N	$42 \le N \le 60$	$61 \le N \le 76$	$77 \le N \le 90$	$91 \le N \le 99$	$100 \le N \le 119$
k(N)	7	$\lfloor 0.13N \rfloor$	$\lfloor 0.12N \rfloor$	10	11
N	$120 \le N \le 129$	$130 \le N \le 139$	$140 \le N \le 166$	$167 \le N \le 177$	$178 \le N \le 188$
k(N)	12	13	14	15	16
N	$189 \le N \le 200$	$201 \le N \le 300$	$300 < N \le 400$		
k(N)	17	$\lceil 0.08 * N \rceil$	$max(24, \lceil 0.07N \rceil)$		

the user of the procedures based on the presented ranking of the multivariate data may use own choice of the critical constant $k(N)$. The choice of k given in Table 2.3.1 is not claimed to be optimal in some sense, but determining $k = k(N)$ makes possible the practical use of the ranking proposed in this section.

Suppose that the mapping (2.3.2) is defined by the formula $\mathbf{u}(\mathbf{z}) = \mathbf{z}$. The ranks

$$R^{(NNt)} = R^{(NNt)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (R_1^{(NNt)}, \ldots, R_n^{(NNt)}) \tag{2.3.23}$$

computed by means of the order statics (2.3.20) by the procedure (2.1.4) are denoted with the superscript (NNt) in order to express that they are based on the nearest neighbors classification and on the dissimilarity measure t. In the formula (2.3.23) sometimes also the notation $R_j^{(NNt)} = R^{(NNt)}(\mathbf{x}_j), j = 1, \ldots, n$ will be used. Now let the mapping (2.3.2)

$$\mathbf{u}(\mathbf{z}) = \begin{cases} \frac{\mathbf{z}}{\|\mathbf{z}\|} & \|\mathbf{z}\| > 0, \\ \mathbf{0}_{d,1} & otherwise. \end{cases} \tag{2.3.24}$$

The ranks

$$R^{(NNst)} = R^{(NNst)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (R_1^{(NNst)}, \ldots, R_n^{(NNst)}) \tag{2.3.25}$$

computed by means of the order statics (2.3.20) by the procedure (2.1.4) are denoted with the superscript $(NNst)$ in order to express that they are based on the nearest neighbors classification, spatial signs (2.3.24) and on the dissimilarity measure t. In the formula (2.3.25) sometimes also the notation $R_j^{(NNst)} = R^{(NNst)}(\mathbf{x}_j), j = 1, \ldots, n$ will be used. We remark that the spatial signs were used for constructing nonparametric tests e.g. by Wang, Peng, and Li (2015), Oja and Randles (2004), Feng and Sun (2016), their use for various purposes can be found in the monograph Oja (2010).

2.4. Multivariate dissimilarity ranks employing an ordering of the rows of random matrix

Since the value of the dissimilarity (used in the previous construction) may depend on the order in which the coordinates of the vector are written, we shall order them in some special way. The resulting ordering (2.4.5) of coordinates is used in Theorem 2.4.1 for construction of multivariate ranks.

For $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{(d,n)}$ let

$$v_j = v_j(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}(\pi_j(\mathbf{x}_i) - \pi_j(\overline{\mathbf{x}})^2, \;\; j = 1, \ldots, d, \quad \overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i. \tag{2.4.1}$$

and

$$\mathbf{M} = \mathbf{M}(\mathbf{x}) = (M_1, \ldots, M_d)^\top, \;\; M_j = median\Big(\pi_j(\mathbf{x}_1), \ldots, \pi_j(\mathbf{x}_n)\Big)' \tag{2.4.2}$$

denote the coordinate-wise medians. Define (cf. (2.3.10))

$$\mathcal{D}_n^{(*sp)} = \{\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{D}_n^{(sp)};\; \textit{the numbers } \pi_1(\mathbf{M}), \ldots, \pi_d(\mathbf{M}) \textit{ are mutually different}\}. \tag{2.4.3}$$

Obviously the relation

$$v_{j_1} \prec v_{j_2} \;\; \textit{if } v_{j_1} < v_{j_2} \;\; \textit{or } v_{j_1} = v_{j_2}\,,\; \pi_{j_1}(\mathbf{M}) < \pi_{j_2}(\mathbf{M}) \tag{2.4.4}$$

is a linear ordering of the numbers $v_1, \ldots, v_d$ provided that $\mathbf{x} \in \mathcal{D}_n^{(*sp)}$. Suppose that $v_{a_1} \prec v_{a_2} \prec \ldots \prec v_{a_d}$ denotes this ordering and

$$s_j = median((|\pi_j(\mathbf{x}_1 - \mathbf{M})|, |\pi_j(\mathbf{x}_2 - \mathbf{M})|, \ldots, |\pi_j(\mathbf{x}_n - \mathbf{M})|)$$

denotes the median absolute deviation of the j-th coordinate. Put

$$\breve{\mathbf{x}} = (\breve{\mathbf{x}}_1, \ldots, \breve{\mathbf{x}}_n), \quad \breve{\mathbf{x}}_i = \Big(\frac{\pi_{a_1}(\mathbf{x}_i - \mathbf{M})}{s_{a_1}}, \ldots, \frac{\pi_{a_d}(\mathbf{x}_i - \mathbf{M})}{s_{a_d}}\Big)^\top, \tag{2.4.5}$$

to obtain standardized observations invariant under coordinate-wise shift and multiplication by a positive constant.

Theorem 2.4.1. *Suppose that t is a dissimilarity measure, $\mathcal{D}_n^{(*sp)}$ is the set (2.4.3) and a sequence of d-dimensional random vectors $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ has a density with respect to the multivariate Lebesgue measure.*

(I) The equality $\mu_L(\mathbb{R}^{(d,n)} - \mathcal{D}_n^{(*sp)}) = 0$ holds and $\mathbf{X} \in \mathcal{D}_n^{(*sp)}$ with probability 1, i.e., the with the notation (2.4.5) the transformation $\breve{\mathbf{X}} = (\breve{\mathbf{X}}_1, \ldots, \breve{\mathbf{X}}_n)$ is well defined with probability 1, and the vector of ranks (cf. (2.3.25), (2.3.23))

$$\begin{aligned}\breve{R}^{(NNst)} &= R^{(NNst)}(\breve{\mathbf{X}}_1, \ldots, \breve{\mathbf{X}}_n) = (\breve{R}_1^{(NNst)}, \ldots, \breve{R}_n^{(NNst)}),\\ \breve{R}^{(NNt)} &= R^{(NNt)}(\breve{\mathbf{X}}_1, \ldots, \breve{\mathbf{X}}_n) = (\breve{R}_1^{(NNt)}, \ldots, \breve{R}_n^{(NNt)})\end{aligned} \tag{2.4.6}$$

are well defined with probability 1.

(II) Suppose that the sequence $(\mathbf{X}_{r_1}, \ldots, \mathbf{X}_{r_n})$ has the same distribution for every permutation $(r_1, \ldots, r_n) \in \mathcal{R}^{(n)}$ and the set $C \subset \mathcal{D}_n^{(*sp)}$ is such that $P(\mathbf{X} \in C) > 0$. If $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in C$ implies that $(\mathbf{x}_{r_1}, \ldots, \mathbf{x}_{r_n}) \in C$ for each permutation $(r_1, \ldots, r_n) \in \mathcal{R}^{(n)}$, then for the ranks (2.4.6) the equalities

$$P(\breve{R}^{(NNst)} = r | \mathbf{X} \in C) = \frac{1}{n!}, \quad P(\breve{R}^{(NNt)} = r | \mathbf{X} \in C) = \frac{1}{n!} \tag{2.4.7}$$

hold for each permutation $r \in \mathcal{R}^{(n)}$.

Proof. (I) The set $\mathbb{R}^{(d,n)} - \mathcal{D}_n^{(sp)}$ has the multivariate Lebesgue measure zero, because it is a subset of $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{(d,n)}$ such that $\sum_{k=1}^{d}\pi_k(\mathbf{x}_{i_1}) = \sum_{k=1}^{d}\pi_k(\mathbf{x}_{i_2})$ for some $1 \le i_1 < i_2 \le n$. Hence the set $\mathbb{R}^{(d,n)} - \mathcal{D}_n^{(*sp)}$ is a subset of the set

$$(\mathbb{R}^{(d,n)} - \mathcal{D}_n^{(sp)}) \cup \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{(d,n)}; \pi_{j_1}(\overline{|\mathbf{x}|}) = \pi_{j_2}(\overline{|\mathbf{x}|}) \textit{ for some } \; 1 \le j_1 < j_2 \le d\},$$

which has the multivariate Lebesgue measure zero. Since $\mathbf{X}$ has a density with respect to the multivariate Lebesgue measure, the assertion (I) is true.

(II) Since the set C is invariant under permutation of the columns of $\mathbf{x}$, for each permutation $r \in \mathcal{R}^{(n)}$

$$P(\breve{R}^{(NNst)} = r, \mathbf{X} \in C) = P(\breve{R}^{(NNst)} = i_d, \mathbf{X} \in C),$$

where i_d is the identical permutation of the set $\{1, \ldots, n\}$. This means that the first equation in (2.4.7) is true, the second one can be proved similarly. □

2.5. Dissimilarities used in this paper

A simple dissimilarity defined on $\mathbb{R}^d$ used in this paper is the Euclidean norm

$$t_1(\mathbf{b}) = \sqrt{\mathbf{b}'\mathbf{b}}. \tag{2.5.1}$$

In the next sections we shall consider for constructing the ranks (2.3.23) or (2.3.25) also another version of dissimilarity. In its computation the inverse

$$arccotg : (-\infty, +\infty) \to (0, \pi) \tag{2.5.2}$$

of the function $cotg(x)$ is used. Suppose that $\mathbf{b} = (b_1, \ldots, b_d)' \in \mathbb{R}^d$ has all components b_j different from zero. Put $b^* = (b_d, b_{d-1}, \ldots, b_1)'$ and define the vectors $\mathbf{v}_0 = \mathbf{v}_0(\mathbf{b})$, $\mathbf{v}_0^* = \mathbf{v}_0(\mathbf{b}^*)$, $\mathbf{v} = \mathbf{v}(\mathbf{b})$ from $\mathbb{R}^{d-1}$ by formula

$$\pi_i(\mathbf{v}_0) = \begin{cases} arccotg\Big(\frac{b_d}{b_{d-1}}\Big) & i = d-1, \\ arccotg\Big(cos(\mathbf{v}_0(i+1))\frac{b_{i+1}}{b_i}\Big) & i = d-2, \ldots, 1, \end{cases}$$

$$\pi_i(\mathbf{v}_0^*) = \begin{cases} arccotg\Big(\frac{b_d^*}{b_{d-1}^*}\Big) & i = d-1, \\ arccotg\Big(cos(\mathbf{v}_0^*(i+1))\frac{b_{i+1}^*}{b_i^*}\Big) & i = d-2, \ldots, 1, \end{cases}$$

$$\pi_i(\mathbf{v}) = max\{|\pi_i(\mathbf{v}_0)|, |\pi_i(\mathbf{v}_0^*)|\}. \tag{2.5.3}$$

Now define the dissimilarity t_2 by the formula

$$t_2 = t_2(\mathbf{b}) = \sqrt{\mathbf{b}'\mathbf{b}(1 + d\mathbf{v}'\mathbf{v})}. \tag{2.5.4}$$

For zero values modify this formula as follows. Let *ind* denote the coordinate of non-zero values of $\mathbf{b}$ and $\mathbf{b}^* = \mathbf{b}(ind)$ be the restriction of $\mathbf{b}$ to the index set *ind* (the subvector of all non-zero components). Put

$$t_2 = t_2(\mathbf{b}) = \begin{cases} t_2(\mathbf{b}^*) & \#(ind) > 1, \\ |b| & \textit{if ind consists of one element,} \\ 0 & \textit{otherwise.} \end{cases} \tag{2.5.5}$$

We remark that the user of the older version of the MATLAB has to define the function (2.5.2) as a new MATLAB function. In the MATLAB an inverse function $acot : (-\infty, +\infty) \to (-\pi/2, \pi/2)$ of the $arccotg$ is bult-in, but $acot(x)$ tends to $-\pi/2$ or $\pi/2$ if x tends to zero from the left or from the right, respectively. Obviously, the function $arccotg(x) = acot(x)$ for $x \geq 0$, and $arccotg(x) = \pi + acot(x)$ otherwise, is the continuous function (2.5.2).

3. Testing the location hypothesis

It is assumed throughout this Section that $\mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,n_i}$ is a random sample from the distribution of the d-dimensional random vector

$$\mathbf{X}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, q, \tag{3.1}$$

where the vector of constants $\boldsymbol{\mu}_i$ denotes the location parameter, the distribution of the column random vector $\boldsymbol{\epsilon}_i$ is the same for all indexes i, possesses a density with respect to the multivariate Lebesgue measure and the random samples are independent. The task is to test the null hypothesis

$$H_0 : \ \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_q \tag{3.2}$$

against the alternative that there exist indexes i_1, i_2 such that $\boldsymbol{\mu}_{i_1} \neq \boldsymbol{\mu}_{i_2}$. Since the topic of this paper is nonparametric testing with possible application to high-dimensional data, we shall consider only nonparametric tests based on statistics defined in the setting not requiring the assumption that the sample size is greater than dimension of the observations.

3.1. Tests in the two-sample case

In this Subsection we assume that $q = 2$, hence the null hypothesis

$$H_0 : \ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2. \tag{3.1.1}$$

We shall construct tests of the null hypothesis (3.1.1) by means of the Wilcoxon test statistic. We recall that if $\mathbf{R} = (R_1, \ldots, R_{n_1}, R_{n_1+1}, \ldots, R_{n_1+n_2})$ is a random vector with values in the set of permutations $\mathcal{R}^{(n_1+n_2)}$, then the Wilcoxon two-sample statistic $S_{n_1,n_2} = \sum_{i=1}^{n_1} R_i$ and the critical region

$$Re(S_{n_1,n_2}, w) = \{\textit{reject the null hypothesis if } S_{n_1,n_2} \leq w \textit{ or } S_{n_1,n_2} \geq n_1(n_1+n_2+1) - w\}, \tag{3.1.2}$$

where $w = w_{\alpha/2}$, $P(\sum_{i=1}^{n_1} R_i \leq w_{\alpha/2}) = \alpha/2$ and this probability is computed under the assumption that the random vector $\mathbf{R}$ is uniformly distributed over the set $\mathcal{R}^{(n_1+n_2)}$. As is well-known, in the usual one-dimensional case the critical region $Re(S_{n_1,n_2}, w_{\alpha/2})$ yields a test of the null hypothesis at the significance level α.

To handle the multivariate case, suppose that $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{n_1+n_2})$ denotes the pooled sample, i.e.,

$$\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{n_1+n_2}) = (\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{1,n_1}, \mathbf{X}_{2,1}, \ldots, \mathbf{X}_{2,n_2}). \tag{3.1.3}$$

Let $\breve{\mathbf{Z}} = (\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2})$ denote transformation (2.4.5) of the pooled random sample (3.1.3). In the following Theorem we use a vector $\mathbf{c} \in \mathbb{R}^d$ which yields large (possibly maximal) values of the Friedman index (4.4). The computation of this vector is described in Section 4.

Theorem 3.1.1. *Put $n = n_1 + n_2$. Let $\mathbf{c} = \mathbf{c}_{FPP}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n)$ be the Friedman projection pursuit vector (4.8). Consider the ordering (2.2.1) of the sequence of vectors $\breve{\mathbf{Z}}$, the order statistics (2.2.2) and in accordance with (2.1.4) put $R_i^{(FPP)} = r_i$ whenever $\breve{\mathbf{Z}}_i = \breve{\mathbf{Z}}^{(r_i)}$. The vector of ranks $R^{(FPP)} = (R_1^{(FPP)}, \ldots, R_{n_1+n_2}^{(FPP)})$ is uniquely defined and takes values in $\mathcal{R}^{(n_1+n_2)}$ with probability 1. If the null hypothesis (3.1.1) holds, then the vector $R^{(FPP)}$ is uniformly distributed over the set $\mathcal{R}^{(n_1+n_2)}$ and the statistic*

$$S_{n_1,n_2}^{(FPP)} = \sum_{i=1}^{n_1} R_i^{(FPP)} \tag{3.1.4}$$

has the same distribution as the two-sample Wilcoxon statistic S_{n_1,n_2} based on sampling from one-dimensional continuous distribution. Hence test with the critical region $Re(S_{n_1,n_2}^{(FPP)}, w_{\alpha/2})$ from (3.1.2) is a test of the null hypothesis at the significance level α.

Proof. As this Theorem can be easily proved by an application of Theorem 2.2.1, the proof is omitted. □

As before let $\breve{\mathbf{Z}} = (\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2})$ denote transformation (2.4.5) of (3.1.3). Then Theorem 2.4.1 implies that the ranks

$$\breve{R}^{(NNst)} = R^{(NNst)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2}) = (\breve{R}_1^{(NNst)}, \ldots, \breve{R}_{n_1+n_2}^{(NNst)}), \tag{3.1.5}$$

$$\breve{R}^{(NNt)} = R^{(NNt)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2}) = (\breve{R}_1^{(NNt)}, \ldots, \breve{R}_{n_1+n_2}^{(NNt)}), \tag{3.1.6}$$

computed by means of a dissimilarity t are uniquely defined almost surely and if the null hypothesis (3.1.1) holds, they are uniformly distributed. As their use in the Wilcoxon test yields in some cases power which is for the ranks $\breve{R}^{(NNst)}$ better than for the ranks $\breve{R}^{(NNt)}$, while in some other cases the power comparison is opposite, we propose to combine these ranks as follows. In this combination we use also the fact that the coordinate-wise median of the matrix $\breve{\mathbf{Z}}$ is the d-dimensional zero vector $\mathbf{0}$.

Theorem 3.1.2. *Compute (cf. (2.3.25), (2.3.23))*

$$\overline{R}^{(NNst)} = R^{(NNst)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2}, \mathbf{M}), \quad \overline{R}^{(NNt)} = R^{(NNt)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2}, \mathbf{M}),$$

where $\mathbf{M} = \mathbf{0}$ *if all vectors* $\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_{n_1+n_2}$ *are different from* $\mathbf{0}$, *and* $\mathbf{M} = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2} \breve{\mathbf{Z}}_i$ *otherwise. Let* $\overline{R}^{(NNst)(1)} < \ldots < \overline{R}^{(NNst)(n_1+n_2)}$ *denote the ordering of the first* $n_1 + n_2$ *coordinates of* $\overline{R}^{(NNst)}$ *according to their magnitude and similarly,* $\overline{R}^{(NNt)(1)} < \ldots < \overline{R}^{(NNt)(n_1+n_2)}$ *denote the ordering of the first* $n_1 + n_2$ *coordinates of* $\overline{R}^{(NNt)}$. *Use the coefficients*

$$c_{NNst} = \sum_{i=1}^{n_1+n_2} \Big(\overline{R}^{(NNst)(j)} - \frac{n_1+n_2+1}{2}\Big)\Big(\overline{R}^{(NNst)(n_1+n_2+1-j)} - \frac{n_1+n_2+1}{2}\Big), \quad (3.1.7)$$

$$c_{NNt} = \sum_{i=1}^{n_1+n_2} \Big(\overline{R}^{(NNt)(j)} - \frac{n_1+n_2+1}{2}\Big)\Big(\overline{R}^{(NNt)(n_1+n_2+1-j)} - \frac{n_1+n_2+1}{2}\Big) \quad (3.1.8)$$

and put (cf. (3.1.5), (3.1.6))

$$\tilde{R}^{(t)} = (\tilde{R}_1^{(t)}, \ldots, \tilde{R}_{n_1+n_2}^{(t)}) = \begin{cases} \breve{R}^{(NNt)} & |c_{NNt}| \geq |c_{NNst}|, \\ \breve{R}^{(NNst)} & \textit{otherwise.} \end{cases} \quad (3.1.9)$$

The test statistic

$$\tilde{S}_{n_1,n_2}^{(t)} = \sum_{i=1}^{n_1} \tilde{R}_i^{(t)} \quad (3.1.10)$$

is uniquely defined with probability 1. *If the null hypothesis (3.1.1) holds, then the vector of ranks (3.1.9) is uniformly distributed over the set of all permutations* $\mathcal{R}^{(n_1+n_2)}$ *and the test statistic (3.1.10) has the same distribution as the Wilcoxon statistic* S_{n_1,n_2} *based on sampling from one-dimensional continuous distribution. Hence the test with the critical region* $Re(\tilde{S}_{n_1,n_2}^{(t)}, w_{\alpha/2})$ *from (3.1.2) is a test of the null hypothesis at the significance level* α.

Proof. The proof follows from Theorem 2.4.1. Indeed, the sets $A_t = \{|c_{NNt}| \geq |c_{NNst}|\}$, $A_{st} = \{|c_{NNt}| < |c_{NNst}|\}$ are invariant under permutation of the columns of the random matrix $\mathbf{Z}$ from (3.1.3). This together with Theorem 2.4.1 means that under the validity of the null hypothesis for each permutation $r \in \mathcal{R}^{(n)}$

$$P(\breve{R}^{(NNt)} = r|A_t) = \frac{1}{n!}, \quad P(\breve{R}^{(NNst)} = r|A_{st}) = \frac{1}{n!},$$

which together with the law of total probability means that the ranks are uniformly distributed, and the Theorem is obviously true. □

The coefficient (3.1.7) can be considered as a measure of asymmetry of the vector $\overline{R}^{(NNst)}$ (the same holds for (3.1.8), $\overline{R}^{(NNt)}$). Thus one may expect that the rule (3.1.9) chooses the type of ranks for which the value of the statistic (3.1.10) is closer to the endpoints of its range, which could under alternative favorably influence the probability of rejection.

The test of the location hypothesis based on the statistic (3.1.10) is exchangeable equivariant and invariant under coordinate-wise shift and multiplication by a positive constant, i.e., its

decision based on $\mathbf{ODX}_{1,1}+\boldsymbol{\kappa},\ldots,\mathbf{ODX}_{1,n_1}+\boldsymbol{\kappa},\mathbf{ODX}_{2,1}+\boldsymbol{\kappa},\ldots,\mathbf{ODX}_{2,n2}+\boldsymbol{\kappa}$ is the same for any permutation matrix $\mathbf{O}$, diagonal matrix $\mathbf{D}$ with a positive diagonal, and any vector $\boldsymbol{\kappa}$.

In the simulation comparisons we shall use the following known tests of (3.1.1). We remark that all these tests are spherically location equivariant, i.e., their decision based on $\mathbf{CX}_{i,j}+\boldsymbol{\kappa}$, $i=1,2$, $j=1,\ldots,n_i$ is the same for any orthogonal matrix $\mathbf{C}$ and any vector $\boldsymbol{\kappa}$.

Let us define the ranking based on the first principal component. Suppose that $\mathbf{C}=\frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}(\mathbf{Z}_i-\overline{\mathbf{Z}})(\mathbf{Z}_i-\overline{\mathbf{Z}})^\top$ denotes the sample covariance matrix of the pooled random sample (3.1.3), λ_1 denote its largest characteristic root and $\tilde{\mathbf{c}}=\tilde{\mathbf{c}}(\mathbf{Z})$ be a non-zero characteristic vector corresponding to λ_1. Let $\mathbf{c}=\mathbf{c}(\mathbf{Z})$ be the vector (2.3.4). Compute the scalar products $\mathbf{c}^\top\mathbf{Z}_1,\ldots,\mathbf{c}^\top\mathbf{Z}_{n_1+n_2}$ and use the ordering (2.2.1) to compute by means of (2.1.4) the ranks $R^{(pc)}=(R_1^{(pc)},\ldots,R_{n_1+n_2}^{(pc)})$ of the vectors $\mathbf{Z}_1,\ldots,\mathbf{Z}_{n_1+n_2}$. According to Theorem 2.2.1 the statistic

$$S_{n_1,n_2}^{(pc)}=\sum_{i=1}^{n_1}R_i^{(pc)} \tag{3.1.11}$$

is uniquely defined almost surely and under the validity of (3.1.1) it has the same distribution as the Wilcoxon statistic S_{n_1,n_2} based on sampling from one-dimensional continuous distribution. This statistic (3.1.11), based on ranking by means of the first principal component, has been proposed for testing (3.1.1) by Rousson (2002). As the Referee suggested, as far as the use of the principal component is concerned, the test procedure could be based also on more than one component. This is a possible matter of further research, and also an implementation of the idea of invariant coordinate selection from Nordhausen, Oja, and Tyler (2006) for constructing invariant multivariate rank tests (in high dimension).

The statistic

$$T_{n_1,n_2}=\frac{1}{n_1(n_1-1)}\sum_{i\neq j}^{n_1}\mathbf{X}_{1,i}'\mathbf{X}_{1,j}+\frac{1}{n_2(n_2-1)}\sum_{i\neq j}^{n_2}\mathbf{X}_{2,i}'\mathbf{X}_{2,j}-\frac{2}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{X}_{1,i}'\mathbf{X}_{2,j}$$

is presented by Chen and Qin (2010). To estimate the variance of this test statistic, put

$$\overline{\mathbf{X}}_{i(j,k)}=\frac{1}{n_i-2}\sum_{j\neq r\neq k}\mathbf{X}_{i,r},\quad \overline{\mathbf{X}}_{i(\ell)}=\frac{1}{n_i-1}\sum_{j\neq\ell}^{n_i}\mathbf{X}_{i,j},$$

$$\widehat{tr(\boldsymbol{\Sigma}_i^2)}=\frac{1}{n_i(n_i-1)}tr\Big(\sum_{j\neq k}^{n_i}(\mathbf{X}_{i,j}-\overline{\mathbf{X}}_{i(j,k)})\mathbf{X}_{i,j}'(\mathbf{X}_{i,k}-\overline{\mathbf{X}}_{i(j,k)})\mathbf{X}_{i,k}'\Big),$$

$$\widehat{tr(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}=\frac{1}{n_1n_2}tr\Big(\sum_{\ell=1}^{n_1}\sum_{k=1}^{n_2}(\mathbf{X}_{1,\ell}-\overline{\mathbf{X}}_{1(\ell)})\mathbf{X}_{1,\ell}'(\mathbf{X}_{2,k}-\overline{\mathbf{X}}_{2(k)})\mathbf{X}_{2,k}'\Big),$$

Under the conditions postulated by Chen and Qin (2010), and under the validity of the null hypothesis the statistic

$$\hat{\sigma}_{n1}^2=\frac{2}{n_1(n_1-1)}\widehat{tr(\boldsymbol{\Sigma}_1^2)}+\frac{2}{n_2(n_2-1)}\widehat{tr(\boldsymbol{\Sigma}_2^2)}+\frac{4}{n_1n_2}\widehat{tr(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}$$

is a ratio-consistent estimate of the variance of T_{n_1,n_2} and the normalized ratio is asymptotically normal, as both n_1, n_2 and d tend to infinity. Therefore Chen and Qin (2010) propose to reject the null hypothesis whenever $\frac{T_{n_1,n_2}}{\sqrt{\hat{\sigma}_{n1}^2}}$ exceeds ξ_α, where ξ_α is the upper quantile of $N(0,1)$. Thus the probability of rejection the null hypothesis (3.1.1) by the test of Chen and Qin is

$$P_{CQ}=P\Big(\frac{T_{n_1,n_2}}{\sqrt{\hat{\sigma}_{n1}^2}}>\xi_\alpha\Big). \tag{3.1.12}$$

Test statistic proposed by Biswas and Ghosh (2014) is defined by means of the formulas

$$
\begin{aligned}
T^*_{n_1,n_2} &= \frac{(n_1+n_2)\hat{\lambda}(1-\hat{\lambda})}{2\hat{\sigma}_0^2} T_{n_1,n_2}, \qquad (3.1.13)\\
\hat{\lambda} &= \frac{n_1}{n_1+n_2}, \quad \hat{\sigma}_0^2 = \frac{n_1 S_1 + n_2 S_2}{n_1+n_2}, \quad T_{n_1,n_2} = \|\widehat{\boldsymbol{\mu}}_F - \widehat{\boldsymbol{\mu}}_G\|^2,\\
S_1 &= \Big[\binom{n_1}{3}^{-1} \sum_{1\le i<j<k\le n_1} \|\mathbf{X}_{1,i}-\mathbf{X}_{1,j}\|\,\|\mathbf{X}_{1,i}-\mathbf{X}_{1,k}\|\Big] - \Big[\binom{n_1}{2}^{-1} \sum_{1\le i<j\le n_1} \|\mathbf{X}_{1,i}-\mathbf{X}_{1,j}\|\Big]^2,\\
S_2 &= \Big[\binom{n_2}{3}^{-1} \sum_{1\le i<j<k\le n_2} \|\mathbf{X}_{2,i}-\mathbf{X}_{2,j}\|\,\|\mathbf{X}_{2,i}-\mathbf{X}_{2,k}\|\Big] - \Big[\binom{n_2}{2}^{-1} \sum_{1\le i<j\le n_2} \|\mathbf{X}_{2,i}-\mathbf{X}_{2,j}\|\Big]^2,\\
\widehat{\boldsymbol{\mu}}_F &= \left[\binom{n_1}{2}^{-1} \sum_{i=1}^{n_1-1}\sum_{j=i+1}^{n_1} \|\mathbf{X}_{1,i}-\mathbf{X}_{1,j}\|,\ (n_1 n_2)^{-1} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \|\mathbf{X}_{1,i}-\mathbf{X}_{2,j}\|\right],\\
\widehat{\boldsymbol{\mu}}_G &= \left[(n_1 n_2)^{-1} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \|\mathbf{X}_{1,i}-\mathbf{X}_{2,j}\|,\ \binom{n_2}{2}^{-1} \sum_{i=1}^{n_2-1}\sum_{j=i+1}^{n_2} \|\mathbf{X}_{2,i}-\mathbf{X}_{2,j}\|\right].
\end{aligned}
$$

The null hypothesis (3.1.1) is rejected if $T^*_{n_1,n_2} > \chi^2_{1-\alpha}$ where $\chi^2_{1-\alpha}$ is $(1-\alpha)th$ quantile of the χ^2 distribution with 1 degree of freedom, this result is derived for distributions with finite second moments. Thus the probability of rejection the null hypothesis (3.1.1) by the test of Biswas and Ghosh is

$$P_{BG} = P(\, T^*_{n_1,n_2} > \chi^2_{1-\alpha}\,). \qquad (3.1.14)$$

Finally, we are going to describe the test statistic proposed by Feng, Zou, and Wang (2016). The basic tool of this statistic are a diagonal matrix $\mathbf{D}_i$ with positive diagonal and a vector $\boldsymbol{\theta}_i \in \mathbb{R}^d$ such that for $i = 1, 2$

$$\frac{1}{n_i}\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}) = \mathbf{0}, \quad \frac{p}{n_i}\mathrm{diag}\Big\{\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij})U(\boldsymbol{\epsilon}_{ij})^\top\Big\} = \mathbf{I}_p, \quad \boldsymbol{\epsilon}_{ij} = \mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i).$$

This estimates are computed by means of the following iterative procedure.

Computation of $\mathbf{D}$ *and* $\boldsymbol{\theta}$.

Step 1. Put $\mathbf{D}_i = diag(v_{i1}, \ldots, v_{ip})$, where v_{ij} is the sample variance of the jth coordinate of the ith sample, and $\boldsymbol{\theta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{X}_{ij}$.

Step 2. Put

$$
\begin{aligned}
\boldsymbol{\epsilon}_{ij} &:= \mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i), \ \ j = 1, \ldots, n_i,\\
\boldsymbol{\theta}_i &:= \boldsymbol{\theta}_i + \frac{\mathbf{D}_i^{1/2}\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij})}{\sum_{j=1}^{n_i}\|\boldsymbol{\epsilon}_{ij}\|^{-1}},\\
\mathbf{D}_i &:= p\mathbf{D}_i^{1/2}\mathrm{diag}\{\frac{1}{n_i}\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}U(\boldsymbol{\epsilon}_{ij})^\top\}\mathbf{D}_i^{1/2}
\end{aligned}
$$

Let δ be a chosen inaccuracy (in the simulations mentioned in this paper $\delta = 0.000001$). The stopping rule is: Repeat Step 2 while

$$\max\Big(\|\frac{1}{n_i}\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}) - \mathbf{0}\|,\ \|\frac{p}{n_i}\mathrm{diag}\Big\{\sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij})U(\boldsymbol{\epsilon}_{ij})^\top\Big\} - \mathbf{I}_p\|\Big) \ge \delta. \qquad (3.1.15)$$

Since in some simulation trials this procedure seemed not to converge and typically the maximal number of iterations is less than 500, in the simulations mentioned in this paper and including these statistics, the stopping rule was modified in such a way that either the number of the iteration is 500 or the inequality (3.1.15) does not hold.

Computation of the test statistic. Let

$$R_n = -\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U^\top\Big(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j})\Big)U\Big(\hat{\mathbf{D}}_{2,j}^{-1/2}(\mathbf{X}_{2j}-\hat{\boldsymbol{\theta}}_{1,i})\Big),$$

where $\hat{\boldsymbol{\theta}}_{i,j}$ a $\hat{\mathbf{D}}_{i,j}^{-1/2}$ are the location vectors and scatter matrices using leave-one-out samples $\{\mathbf{X}_{ik}\}_{k\neq j}$. Feng *et al.* (2016) estimate the asymptotic variance by the estimator

$$\sigma_n^2 = \frac{2}{n_1(n_1-1)p^2}\widehat{\mathrm{tr}(\mathbf{A}_1^2)} + \frac{2}{n_2(n_2-1)p^2}\widehat{\mathrm{tr}(\mathbf{A}_2^2)} + \frac{4}{n_1 n_2 p^2}\widehat{\mathrm{tr}(\mathbf{A}_3^\top\mathbf{A}_3)},$$

$$\begin{aligned}
\widehat{\mathrm{tr}(\mathbf{A}_1^2)} &= \frac{p^2\hat{c}_2^2\hat{c}_1^{-2}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}\Big(\tilde{\mathbf{U}}_{1l}^\top\hat{\mathbf{D}}_2^{-1/2}\hat{\mathbf{D}}_1^{1/2}\tilde{\mathbf{U}}_{1k}\Big)^2,\\
\widehat{\mathrm{tr}(\mathbf{A}_2^2)} &= \frac{p^2\hat{c}_1^2\hat{c}_2^{-2}}{n_2(n_2-1)}\sum_{k=1}^{n_2}\sum_{l\neq k}\Big(\tilde{\mathbf{U}}_{2l}^\top\hat{\mathbf{D}}_1^{-1/2}\hat{\mathbf{D}}_2^{1/2}\tilde{\mathbf{U}}_{2k}\Big)^2,\\
\widehat{\mathrm{tr}(\mathbf{A}_3^\top\mathbf{A}_3)} &= \frac{p^2}{n_1 n_2}\sum_{l=1}^{n_1}\sum_{k=1}^{n_2}\Big(\tilde{\mathbf{U}}_{1l}^\top\tilde{\mathbf{U}}_{2k}\Big)^2,
\end{aligned}$$

where $\hat{c}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\|\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{i,j}-\hat{\boldsymbol{\theta}}_{i,j})\|$ and $\tilde{\mathbf{U}}_{ij} = U\Big(\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{ij}-\hat{\boldsymbol{\theta}}_{i,j})\Big)$. Under the conditions imposed by Feng *et al.* (2016) the distribution of the statistic $\frac{R_n}{\sigma_n}$ is asymptotically normal and the probability of rejection the hypothesis (3.1.1) by the test of Feng, Zou and Wang is

$$P_{FZW} = P(\frac{R_n}{\sigma_n} > \xi_\alpha). \tag{3.1.16}$$

3.2. Simulation results in the two-sample case

Unless it is not said otherwise, in the simulations we shall use the matrix

$$\mathbf{V}_{d,m} = diag(\mathbf{dgn}) + \Big(\mathbf{1}_{d,d} - \mathbf{I}_d\Big)\frac{m}{d}, \tag{3.2.1}$$

where $\mathbf{dgn}(j) = 1.5 + \frac{j-1}{d}1.5$, $j = 1,\ldots,d$, $\mathbf{1}_{d,d}$ is the $d\times d$ matrix having on all positions the number 1 and $\mathbf{I}_d$ is the usual identity matrix. For large values of the dimension d we consider in this section the location parameter and the scatter (matrix) parameter

$$\begin{gathered}
\boldsymbol{\mu} = c\,(-\mathbf{1}_{a_1},\mathbf{1}_{b_1})^\top \quad a_1 = \lfloor\tfrac{d}{2}\rfloor,\ b_1 = d-a_1, \qquad \boldsymbol{\mu} = c\,\mathbf{1}_d^\top, \qquad \boldsymbol{\mu} = c\,(1,\mathbf{0}_{1,d-1})^\top\\
\mathbf{V} = \mathbf{I}_d + 0.2\Big(\mathbf{1}_{d,d} - \mathbf{I}_d\Big),
\end{gathered} \tag{3.2.2}$$

where $\mathbf{1}_a = (1,\ldots,1)$ denotes this a-dimensional row vector and $\mathbf{0}_{1,d-1}$ is the $d-1$-dimensional vector of zeros. In some explicitly mentioned cases we shall use the location and scatter parameters

$$\tilde{\boldsymbol{\mu}} = (0, 3*\mathbf{1}_{d-1})^\top, \qquad \tilde{\mathbf{V}} = diag(100,\mathbf{1}_{d-1}). \tag{3.2.3}$$

Let $\mathbf{W}$ be a symmetric positive definite $d\times d$ matrix. By the d-dimensional Cauchy distribution $C_d(\boldsymbol{\mu},\mathbf{W})$ we shall understand the distribution of the random vector $\boldsymbol{\mu}+\mathbf{W}^{1/2}\boldsymbol{\epsilon}^*$, where the coordinates of the d-dimensional random vector $\boldsymbol{\epsilon}^*$ are independent and each has the Cauchy distribution with location 0 and the scale 1. By the d-dimensional chi-square distribution $\chi^2_{d,fre}(\boldsymbol{\mu},\mathbf{W})$ with fre degrees of freedom we shall understand the distribution of the random vector $\boldsymbol{\mu}+\mathbf{W}^{1/2}\boldsymbol{\epsilon}^*$, where the coordinates of the d-dimensional random vector $\boldsymbol{\epsilon}^*$ are independent and each has the chi-square distribution with fre degrees of freedom. By the d-dimensional generalized Pareto distribution $G_d(\boldsymbol{\mu},\mathbf{W},k,\sigma,\theta)$, where $k>0$, $\sigma>0$, θ are

real numbers, we shall understand the distribution of the random vector $\boldsymbol{\mu} + \mathbf{W}^{1/2}\boldsymbol{\epsilon}^*$, where the coordinates od the d-dimensional random vector $\boldsymbol{\epsilon}^*$ are independent and each of them has the density $f(x,k,\sigma,\theta) = (1/\sigma)\Big(1 + k\frac{(x-\theta)}{\sigma}\Big)^{-(1+\frac{1}{k})}$ for $x > \theta$.

The simulation estimates of probabilities of rejection at the significance level $\alpha = 0.05$ are typed with the hat symbol. In all simulation in the two sample scheme we use $n_1 = 10$, $n_2 = 15$, $w_{\alpha/2} = 94$, because for these sample sizes under the validity of the null hypothesis $P(\sum_{j=1}^{n_1} R_j \le 94) = 0.0238$, $P(\sum_{j=1}^{n_1} R_j \le 95) = 0.0273$. Unless it is said otherwise, all the simulations in this Section are based on $N = 3500$ trials.

By $\hat{P}_{FPP}$ we understand the simulation estimate of the probability $P(Re(S^{(FPP)}_{n_1,n_2}, w_{\alpha/2}))$ of rejection of H_0 by the statistic (3.1.4), here we use the notation (3.1.2). By $\hat{P}_t$ we understand the simulation estimate of probability $P(Re(\tilde{S}^{(t)}_{n_1,n_2}, w_{\alpha/2}))$ of rejection H_0 by the statistic (3.1.10), here we use the dissimilarity (2.5.1) or (2.5.4). By $\hat{P}_{Rou}$ we understand the simulation estimate (cf. (3.1.11)) of $P(Re(S^{(pc)}_{n_1,n_2}, w_{\alpha/2}))$, $\hat{P}_{CQ}$ is the simulation estimate of (3.1.12), $\hat{P}_{BG}$ denotes the simulation estimate of (3.1.14) and $\hat{P}_{FZW}$ denotes the simulation estimate of (3.1.16).

The simulation results yielding (under the particular alternative) the largest power will be typed in bold (two methods yielding the difference of the power not exceeding 0.06 will be regarded as equivalent). The estimates of the probability of the first kind of error, obtained under the validity of the null hypothesis and larger than 0.087, will be typed in bold and the concerned method will not be included into the competition for the largest power for the given type of the distribution. First we are dealing with the observations having the dimension $d = 10$.

$$\mathbf{X}_1 \sim N_{10}(0, \mathbf{V}_{10,m_0}),\ \mathbf{X}_2 \sim N_{10}(\boldsymbol{\mu}, \mathbf{V}_{10,m_0}),\ m_0 = 13$$

Table 3.2.1. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$

c	0	1	1.8	4.5
$\hat{P}_{t_2}$	0.046	0.348	0.856	**1**
$\hat{P}_{FPP}$	0.044	0.169	0.634	**1**
$\hat{P}_{CQ}$	0.086	**0.868**	**1**	**1**
$\hat{P}_{Rou}$	0.046	0.106	0.492	**1**
$\hat{P}_{BG}$	0.079	0.250	0.601	0.623
$\hat{P}_{FZW}$	0.064	**0.918**	**1**	**1**
$\hat{P}_{t_1}$	0.045	0.113	0.693	**1**

Table 3.2.2. $\boldsymbol{\mu} = c\mathbf{1}_{10}^\top$.

c	0.8	1.1	1.6
$\hat{P}_{t_2}$	0.255	0.430	0.715
$\hat{P}_{FPP}$	0.330	0.556	0.861
$\hat{P}_{CQ}$	**0.477**	**0.709**	**0.937**
$\hat{P}_{Rou}$	0.343	0.575	0.876
$\hat{P}_{BG}$	0.158	0.233	0.412
$\hat{P}_{FZW}$	0.381	0.602	**0.882**
$\hat{P}_{t_1}$	0.383	0.558	0.862

Table 3.2.3. $\boldsymbol{\mu} = c\,(1, \mathbf{0}_{1,d-1})^\top$.

c	1	1.5
$\hat{P}_{t_2}$	0.371	0.668
$\hat{P}_{FPP}$	0.483	0.819
$\hat{P}_{CQ}$	**0.631**	**0.911**
$\hat{P}_{Rou}$	0.489	0.834
$\hat{P}_{BG}$	0.202	0.375
$\hat{P}_{FZW}$	0.525	0.843
$\hat{P}_{t_1}$	0.477	0.819

$\mathbf{X}_1 \sim C_{10}(0, \mathbf{V}_{10,m_0})$, $\mathbf{X}_2 \sim C_{10}(\boldsymbol{\mu}, \mathbf{V}_{10,m_0})$, $m_0 = 13$.

Table 3.2.4. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$.

c	0	5	12	19
$\hat{P}_{t_2}$	0.045	0.517	**0.898**	**0.931**
$\hat{P}_{FPP}$	0.047	0.207	**0.868**	**0.972**
$\hat{P}_{CQ}$	**0.097**	0.225	0.586	0.767
$\hat{P}_{Rou}$	0.047	0.073	0.233	0.460
$\hat{P}_{BG}$	**0.112**	0.127	0.225	0.337
$\hat{P}_{FZW}$	0.029	**0.603**	**0.869**	0.904
$\hat{P}_{t_1}$	0.049	0.148	0.731	0.870

Table 3.2.5. $\boldsymbol{\mu} = c\mathbf{1}_{10}^\top$.

c	5	9	13
$\hat{P}_{t_2}$	**0.548**	**0.855**	**0.942**
$\hat{P}_{FPP}$	0.347	0.666	0.841
$\hat{P}_{CQ}$	0.247	0.419	0.557
$\hat{P}_{Rou}$	0.344	0.665	0.843
$\hat{P}_{BG}$	0.121	0.160	0.219
$\hat{P}_{FZW}$	0.393	0.724	0.832
$\hat{P}_{t_1}$	0.463	0.763	**0.889**

Table 3.2.6. $\boldsymbol{\mu} = c\,(1, \mathbf{0}_{1,d-1})^\top$.

c	5	12
$\hat{P}_{t_2}$	**0.548**	**0.929**
$\hat{P}_{FPP}$	0.347	0.807
$\hat{P}_{CQ}$	0.247	0.526
$\hat{P}_{Rou}$	0.344	0.812
$\hat{P}_{BG}$	0.121	0.205
$\hat{P}_{FZW}$	0.393	0.819
$\hat{P}_{t_1}$	0.463	**0.869**

$\mathbf{X}_1 \sim \chi^2_{10,5}(0, \mathbf{V}_{10,m_0})$, $\mathbf{X}_2 \sim \chi^2_{10,5}(\boldsymbol{\mu}, \mathbf{V}_{10,m_0})$, $m_0 = 13$.

Table 3.2.7. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$.

c	0	1.9	2.5	3.1
$\hat{P}_{t_2}$	0.040	0.149	0.247	0.359
$\hat{P}_{FPP}$	0.043	0.069	0.104	0.164
$\hat{P}_{CQ}$	0.039	0.132	0.304	0.608
$\hat{P}_{Rou}$	0.043	0.056	0.075	0.103
$\hat{P}_{BG}$	0.087	0.116	0.156	0.232
$\hat{P}_{FZW}$	0.056	**0.360**	**0.686**	**0.913**
$\hat{P}_{t_1}$	0.039	0.055	0.074	0.108

Table 3.2.8. $\boldsymbol{\mu} = c\mathbf{1}_{10}^\top$.

c	2	3	5
$\hat{P}_{t_2}$	0.194	0.373	0.730
$\hat{P}_{FPP}$	**0.221**	**0.439**	**0.846**
$\hat{P}_{CQ}$	**0.237**	**0.454**	**0.858**
$\hat{P}_{Rou}$	**0.231**	**0.454**	**0.856**
$\hat{P}_{BG}$	0.118	0.179	0.383
$\hat{P}_{FZW}$	**0.267**	**0.491**	**0.875**
$\hat{P}_{t_1}$	**0.222**	**0.445**	**0.844**

Table 3.2.9. $\boldsymbol{\mu} = c\,(1, \mathbf{0}_{1,d-1})^\top$.

c	3	4.5
$\hat{P}_{t_2}$	0.373	0.641
$\hat{P}_{FPP}$	**0.439**	**0.766**
$\hat{P}_{CQ}$	**0.454**	**0.781**
$\hat{P}_{Rou}$	**0.454**	**0.782**
$\hat{P}_{BG}$	0.179	0.334
$\hat{P}_{FZW}$	**0.491**	**0.807**
$\hat{P}_{t_1}$	**0.445**	**0.763**

$\mathbf{X}_1 \sim G_{10}(0, \mathbf{V}_{10,m_0}, 1, 1, 1)$, $\mathbf{X}_2 \sim G_{10}(\boldsymbol{\mu}, \mathbf{V}_{10,m_0}, 1, 1, 1)$, $m_0 = 13$.

Table 3.2.10. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$.

c	0	2	5	7
$\hat{P}_{t_2}$	0.053	**0.095**	**0.334**	**0.511**
$\hat{P}_{FPP}$	0.047	**0.043**	0.083	0.183
$\hat{P}_{CQ}$	**0.123**	0.129	0.169	0.227
$\hat{P}_{Rou}$	0.049	**0.047**	0.057	0.072
$\hat{P}_{BG}$	**0.119**	0.119	0.124	0.131
$\hat{P}_{FZW}$	0.015	0.029	0.231	0.438
$\hat{P}_{t_1}$	0.046	**0.041**	0.038	0.072

Table 3.2.11. $\boldsymbol{\mu} = c\mathbf{1}_{10}^\top$.

c	6	12	30
$\hat{P}_{t_2}$	**0.352**	**0.671**	**0.936**
$\hat{P}_{FPP}$	**0.365**	**0.666**	**0.917**
$\hat{P}_{CQ}$	0.274	0.424	0.686
$\hat{P}_{Rou}$	**0.362**	**0.673**	**0.932**
$\hat{P}_{BG}$	0.124	0.156	0.339
$\hat{P}_{FZW}$	0.250	0.602	0.840
$\hat{P}_{t_1}$	**0.388**	**0.715**	**0.953**

Table 3.2.12. $\boldsymbol{\mu} = c\,(1, \mathbf{0}_{1,d-1})^\top$.

c	5	7
$\hat{P}_{t_2}$	**0.283**	**0.412**
$\hat{P}_{FPP}$	**0.298**	**0.427**
$\hat{P}_{CQ}$	0.241	0.297
$\hat{P}_{Rou}$	**0.295**	**0.422**
$\hat{P}_{BG}$	0.124	0.128
$\hat{P}_{FZW}$	0.181	0.327
$\hat{P}_{t_1}$	**0.313**	**0.453**

Table 3.2.13. Sampling with $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \tilde{\boldsymbol{\mu}}$, $\tilde{\mathbf{V}}$ is the alternative (3.2.3), N=1000.

	$\mathbf{X}_i \sim N_{10}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim C_{10}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim \chi^2_{10,5}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim G_{10}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}}, 1, 1, 1)$
$\hat{P}_{t_2}$	**1**	0.637	0.908	**0.455**
$\hat{P}_{FPP}$	**0.976**	0.574	0.803	0.367
$\hat{P}_{CQ}$	**0.961**	0.145	0.121	0.160
$\hat{P}_{Rou}$	0.123	0.219	0.053	0.231
$\hat{P}_{BG}$	0.422	0.109	0.106	0.121
$\hat{P}_{FZW}$	**1**	0.437	**0.999**	0.178
$\hat{P}_{t_1}$	**1**	**0.854**	0.931	**0.441**

In the following simulations the dimension of random vectors $d = 1000$ and the scatter parameter is the matrix from (3.2.2). We remark that this matrix $\mathbf{V}$ is used in simulations presented by Wang *et al.* (2015). The number of trials is in this cases $N = 1000$.

$$\mathbf{X}_1 \sim N_{1000}(0, \mathbf{V}),\ \mathbf{X}_2 \sim N_{1000}(\boldsymbol{\mu}, \mathbf{V}).$$

Table 3.2.14. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$, *N=1000.*

c	0	0.25	0.65	0.9
$\hat{P}_{t_2}$	0.058	0.069	0.472	0.933
$\hat{P}_{FPP}$	0.047	0.114	0.524	0.732
$\hat{P}_{CQ}$	**0.091**	0.363	1	1
$\hat{P}_{Rou}$	0.056	0.079	0.363	0.865
$\hat{P}_{BG}$	0.031	0.048	0.380	0.526
$\hat{P}_{FZW}$	0.068	**0.293**	**1**	**1**
$\hat{P}_{t_1}$	0.049	0.071	0.474	0.930

Table 3.2.15. $\boldsymbol{\mu} = c\mathbf{1}^\top_{1000}$, *N=1000.*

c	0.25	0.5	0.6
$\hat{P}_{t_2}$	**0.252**	**0.740**	**0.854**
$\hat{P}_{FPP}$	0.176	0.531	0.629
$\hat{P}_{CQ}$	0.380	0.835	0.930
$\hat{P}_{Rou}$	**0.255**	0.726	**0.865**
$\hat{P}_{BG}$	0.058	0.175	0.271
$\hat{P}_{FZW}$	**0.304**	**0.787**	**0.903**
$\hat{P}_{t_1}$	**0.249**	**0.739**	**0.860**

Table 3.2.16. $\boldsymbol{\mu} = c(1, \mathbf{0}_{1,d-1})^\top$, *N=1000.*

c	0.3	0.5
$\hat{P}_{t_2}$	0.347	**0.740**
$\hat{P}_{FPP}$	0.221	0.517
$\hat{P}_{CQ}$	0.496	0.835
$\hat{P}_{Rou}$	**0.354**	0.726
$\hat{P}_{BG}$	0.077	0.175
$\hat{P}_{FZW}$	**0.411**	**0.787**
$\hat{P}_{t_1}$	0.344	**0.739**

$$\mathbf{X}_1 \sim C_{1000}(0, \mathbf{V}),\ \mathbf{X}_2 \sim C_{1000}(\boldsymbol{\mu}, \mathbf{V}).$$

Table 3.2.17. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$, *N=1000.*

c	0	10	30	100
$\hat{P}_{t_2}$	0.049	**0.047**	**0.796**	**0.988**
$\hat{P}_{FPP}$	0.034	**0.068**	0.508	0.872
$\hat{P}_{CQ}$	0.085	**0.104**	0.372	0.851
$\hat{P}_{Rou}$	0.039	0.039	0.072	0.448
$\hat{P}_{BG}$	**0.114**	0.114	0.116	0.247
$\hat{P}_{FZW}$	0.019	**0.077**	**0.772**	**0.928**
$\hat{P}_{t_1}$	0.049	0.034	0.673	**0.962**

Table 3.2.18. $\boldsymbol{\mu} = c\mathbf{1}^\top_{1000}$, *N=1000.*

c	20	40	60
$\hat{P}_{t_2}$	**0.713**	**0.970**	**0.990**
$\hat{P}_{FPP}$	0.341	0.730	0.880
$\hat{P}_{CQ}$	0.253	0.496	0.625
$\hat{P}_{Rou}$	0.352	0.745	0.886
$\hat{P}_{BG}$	0.114	0.126	0.159
$\hat{P}_{FZW}$	0.400	0.790	0.887
$\hat{P}_{t_1}$	**0.669**	**0.939**	**0.973**

Table 3.2.19. $\boldsymbol{\mu} = c(1, \mathbf{0}_{1,d-1})^\top$, *N=1000.*

c	18	30
$\hat{P}_{t_2}$	**0.638**	**0.919**
$\hat{P}_{FPP}$	0.289	0.564
$\hat{P}_{CQ}$	0.216	0.364
$\hat{P}_{Rou}$	0.299	0.573
$\hat{P}_{BG}$	0.114	0.116
$\hat{P}_{FZW}$	0.327	0.684
$\hat{P}_{t_1}$	**0.583**	**0.872**

$\mathbf{X}_1 \sim \chi^2_{1000,5}(0, \mathbf{V})$, $\mathbf{X}_2 \sim \chi^2_{1000,5}(\boldsymbol{\mu}, \mathbf{V})$.

Table 3.2.20. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$, *N=1000.*

c	0	0.5	1.8	2.7
$\hat{P}_{t_2}$	0.046	0.050	0.199	0.781
$\hat{P}_{FPP}$	0.050	0.048	0.298	0.561
$\hat{P}_{CQ}$	0.000	0.000	0.002	0.206
$\hat{P}_{Rou}$	0.045	**0.058**	0.241	0.772
$\hat{P}_{BG}$	0.040	0.045	0.252	0.508
$\hat{P}_{FZW}$	0.062	**0.111**	**1**	**1**
$\hat{P}_{t_1}$	0.044	**0.051**	0.200	0.771

Table 3.2.21. $\boldsymbol{\mu} = c\mathbf{1}^\top_{1000}$, *N=1000.*

c	1	100	300
$\hat{P}_{t_2}$	0.332	**1**	**1**
$\hat{P}_{FPP}$	0.245	**0.999**	**1**
$\hat{P}_{CQ}$	0.001	**1**	**1**
$\hat{P}_{Rou}$	0.338	**1**	**1**
$\hat{P}_{BG}$	0.087	0.520	0.520
$\hat{P}_{FZW}$	**0.422**	**1**	**1**
$\hat{P}_{t_1}$	0.331	**1**	**1**

Table 3.2.22. $\boldsymbol{\mu} = c(1, \mathbf{0}_{1,d-1})^\top$, *N=1000.*

c	0.8	1.3
$\hat{P}_{t_2}$	**0.231**	0.527
$\hat{P}_{FPP}$	0.189	0.433
$\hat{P}_{CQ}$	0.001	0.006
$\hat{P}_{Rou}$	**0.242**	0.545
$\hat{P}_{BG}$	0.067	0.127
$\hat{P}_{FZW}$	**0.291**	**0.625**
$\hat{P}_{t_1}$	**0.232**	0.533

$\mathbf{X}_1 \sim G_{1000}(0, \mathbf{V}, 1, 1, 1)$, $\mathbf{X}_2 \sim G_{1000}(\boldsymbol{\mu}, \mathbf{V}, 1, 1, 1)$.

Table 3.2.23. $\boldsymbol{\mu} = c(-\mathbf{1}_{a_1}, \mathbf{1}_{b_1})^\top$, *N=1000.*

c	0	25	50	60
$\hat{P}_{t_2}$	0.048	0.032	0.403	0.629
$\hat{P}_{FPP}$	0.037	**0.130**	0.443	0.589
$\hat{P}_{CQ}$	**0.101**	0.158	0.411	0.510
$\hat{P}_{Rou}$	0.045	0.053	0.082	0.098
$\hat{P}_{BG}$	**0.112**	0.112	0.124	0.131
$\hat{P}_{FZW}$	0.040	**0.156**	**0.629**	**0.750**
$\hat{P}_{t_1}$	0.048	0.027	0.303	0.543

Table 3.2.24. $\boldsymbol{\mu} = c\mathbf{1}^\top_d$, *N=1000.*

c	20	35	40
$\hat{P}_{t_2}$	**0.203**	**0.454**	**0.543**
$\hat{P}_{FPP}$	**0.186**	0.417	0.493
$\hat{P}_{CQ}$	0.193	0.315	0.360
$\hat{P}_{Rou}$	**0.205**	**0.454**	**0.530**
$\hat{P}_{BG}$	0.112	0.113	0.115
$\hat{P}_{FZW}$	0.119	0.311	0.391
$\hat{P}_{t_1}$	**0.229**	**0.492**	**0.575**

Table 3.2.25. $\boldsymbol{\mu} = c(1, \mathbf{0}_{1,d-1})^\top$, *N=1000.*

c	30	45
$\hat{P}_{t_2}$	**0.372**	**0.619**
$\hat{P}_{FPP}$	0.328	0.559
$\hat{P}_{CQ}$	0.279	0.394
$\hat{P}_{Rou}$	**0.378**	0.587
$\hat{P}_{BG}$	0.113	0.118
$\hat{P}_{FZW}$	0.237	0.458
$\hat{P}_{t_1}$	**0.406**	**0.653**

Table 3.2.26. Sampling with $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \tilde{\boldsymbol{\mu}}$, $\tilde{\mathbf{V}}$ is the alternative (3.2.3). N=1000.

	$\mathbf{X}_i \sim N_{1000}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim C_{1000}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim \chi^2_{1000,5}\boldsymbol{\mu}_i, \tilde{\mathbf{V}})$	$\mathbf{X}_i \sim G_{1000}(\boldsymbol{\mu}_i, \tilde{\mathbf{V}}, 1, 1, 1)$
$\hat{P}_{t_2}$	**1**	0.149	**1**	0.040
$\hat{P}_{FPP}$	**1**	**0.550**	**0.953**	**0.485**
$\hat{P}_{CQ}$	**1**	0.204	**1**	0.202
$\hat{P}_{Rou}$	**1**	0.227	**1**	0.154
$\hat{P}_{BG}$	0.519	0.119	0.499	0.108
$\hat{P}_{FZW}$	**1**	0	**1**	0
$\hat{P}_{t_1}$	**1**	0.434	**1**	0.054

Before summarizing these simulation results, we remark that the test of Biswas and Ghosh, the test of Chen and Qin, and the test of Feng, Zou and Wang, use rejection regions based on the asymptotic distribution of the underlying statistic and do not possess null distribution under the validity of the null hypothesis, they attain the chosen significance level only asymptotically (as the sample sizes tend to infinity). In contradistinction to this, the test of Rousson, the test based on the Wilcoxon statistic and a chosen dissimilarity, and the test based on the Wilcoxon statistic and the sample Friedman projection pursuit vector, possess null distribution under the validity of the null hypothesis, and by a suitable choice of the rejection constant one can achieve a wished finite sample significance level. This does not contradict with the fact that under validity of H_0 the simulation estimates of $\hat{P}_{t_2}$ vary a little with the type of the sampled continuous distributions, which is caused by the fact that simulations are not absolutely prefect (the same holds for $\hat{P}_{FPP}$, $\hat{P}_{Rou}$ and $\hat{P}_{t_1}$).

First pay attention to the previously published tests. It can be seen from the tables that the test of Biswas and Ghosh does not yield for the considered sample sizes and types of distribution the power belonging to the strongest category.

Even though for the dimension of the data $d = 10$ the test of Chen and Qin yields often the power belonging to the strongest category (in Tables 3.2.1, 3.2.2, 3.2.3, 3.2.8, 3.2.9 and 3.2.13), for Cauchy distribution its size (i.e., the probability of the first kind error) equals 0.097 and for the generalized Pareto 0.123 (Tables 3.2.4 and 3.2.10), which considerably exceeds the nominal level 0.05. For $d = 1000$ the test of Chen and Qin has for normal distribution size 0.091, for generalized Pareto distribution 0.101 (Tables 3.2.14 and 3.2.23), and on the other hand, in Table 3.2.22 its power is less than 0.01 (considerably less than for the other mentioned tests). Hence the test of Chen and Qin, whose rejection region is based on the asymptotic distribution, requires larger sizes than those considered in the presented simulations.

The test of Feng, Zou and Wang uses estimates based on an iterative procedure. As remarked by Feng and Sun (2016) on p. 2422, even though there is no proof for the existence of these estimates, they seem to work in practice. In the simulations presented in this paper in some rare cases the procedure seemed not to converge in a reasonable time. Therefore the simulation estimates were computed with the inaccuracy $\delta = 0.000001$ in (3.1.15) and the algorithm was instructed to stop either when (3.1.15) does not hold or when the number of iterations reaches 500. This large number of iterations occurred in sampling from 1000 dimensional Cauchy $C_{1000}(\boldsymbol{\mu}, \mathbf{V})$ distribution or from generalized Pareto $G_{1000}(\boldsymbol{\mu}, \mathbf{V}, 1, 1, 1)$ distribution, where the relative frequency of these samples was $150/1000 = 0.15$. The test of Feng, Zou and Wang, although based on the asymptotic distribution, for the considered distributions and sample sizes yields the largest size 0.068 (Table 3.2.14), which may be regarded as an acceptable value.

The power of the tests can be judged from the point of view of the strongest category recorded in the previous simulations.

Table 3.2.27. The number of cases with power of the first category.

	dimension d=10	dimension d=1000
$\hat{P}_{t_2}$	18	22
$\hat{P}_{FPP}$	15	9
$\hat{P}_{Rou}$	12	13
$\hat{P}_{FZW}$	16	24
$\hat{P}_{t_1}$	17	21

First consider the case when the dimension of observations is not too large and close to sample sizes (here d=10, $n_1 = 10$, $n_2 = 15$). The simulations suggest that the performance of the test based on the Wilcoxon statistic (and the dissimilarity t_1 or the dissimilarity t_2) and the performance of the test of Feng, Zou and Wang, appear to be equivalent, because the numbers of the recorded first category power 18–16 do not differ strikingly. When the numerical size of the recorded power is taken as the sole criterion and two methods yielding the difference of the power not exceeding 0.06 are regarded as equivalent, one obtains a similar result. The i-th row of the following table corresponding to the probability $\hat{P}_i$ shows how many times $\hat{P}_i$ is better than the probability in the jth column (e.g, $\hat{P}_{t_2}$ is better than $\hat{P}_{FZW}$ in 15 cases and $\hat{P}_{FZW}$ is better than $\hat{P}_{t_2}$ in 17 cases).

	$\hat{P}_{t_2}$	$\hat{P}_{FPP}$	$\hat{P}_{FZW}$	$\hat{P}_{t_1}$
$\hat{P}_{t_2}$	0	16	15	13
$\hat{P}_{FPP}$	9	0	7	3
$\hat{P}_{FZW}$	17	9	0	10
$\hat{P}_{t_1}$	10	7	9	0

Hence from this point of view $\hat{P}_{FZW}$ appears to have the best ranking. From the point of view of the sampled distribution, the best results for sampling from the normal or from the chi-square distribution yields the test Feng, Zou and Wang, for sampling from the Cauchy distribution or from the generalized Pareto distribution yields the test based on t_2.

For the dimension of observations considerably greater than the sample sizes (here d=1000, $n_1 = 10$, $n_2 = 15$), two competitors appear as the most successful, because the performance of the test based on the Wilcoxon statistic and the dissimilarity t_2 (22 first category cases), and the performance of the test of Feng, Zou and Wang (24 first category cases), appear to be equivalent. When the numerical size of the recorded power is taken as the sole criterion and two methods yielding the difference of the power not exceeding 0.06 are regarded as equivalent, then ranking of these two methods yields similar results (t_2 is better than the test of Feng, Zou and Wang in 12 cases, and FZW is better than t_2 in 12 cases), but the results when FZW test is better than t_2 test are sometimes more striking than in the opposite case (Tables 3.2.14 and 3.2.20). Also in this case when the dimension of observations $d = 1000$, the test based on t_2 is better for sampling from the Cauchy distribution and from the generalized Pareto distribution, the test Feng, Zou and Wang is better for sampling from the normal distribution and from the chi-square distribution. However, when one wants to have a safeguard against the case when the scale parameter of some coordinate of observations strikingly exceeds the scale parameters of the other coordinates (like in (3.2.3)), then the results from Table 3.2.26 suggest that for high dimensions (like d=1000) the use of the test based on the Friedman projection pursuit vector is recommendable.

3.3. Tests of the location hypothesis in the multisample case

Now we are going to deal with testing of the hypothesis (3.2) when in (3.1) the number of the samples $q > 2$. In this Subsection $n = n_1 + n_2 + \ldots + n_q$ denotes the total sample size and

$$\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n) = (\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{1,n_1}, \mathbf{X}_{2,1}, \ldots, \mathbf{X}_{2,n_2}, \ldots, \mathbf{X}_{q,1}, \ldots, \mathbf{X}_{q,n_q}) \tag{3.3.1}$$

the pooled random sample. Suppose that

$$R = (R_{1,1}, \ldots, R_{1,n_1}, R_{2,1}, \ldots, R_{2,n_2}, \ldots, R_{q,1}, \ldots, R_{q,n_q}) \tag{3.3.2}$$

is a random vector with values in the set $\mathcal{R}^{(n)}$ of all permutations of $\{1, \ldots, n\}$. Let us recall that in accordance with Theorem 1, p. 205 of Hájek *et al.* (1999), and the results of Kruskal and Wallis (1952), the Kruskal-Wallis test statistic and its rejection region are

$$Q_{n_1,\ldots,n_q}(R) = \frac{12}{n(n+1)} \sum_{i=1}^{q} n_i \Big(\frac{S_{[i]}(R)}{n_i} - \frac{n+1}{2} \Big)^2, \quad S_{[i]}(R) = \sum_{j=1}^{n_i} R_{i,j}, \ \ i = 1, \ldots, q, \tag{3.3.3}$$

$$Re(Q_{n_1,\ldots,n_q}(R)) = \{\textit{reject the null hypothesis if } Q_{n_1,\ldots,n_q}(R) > \chi^2_{q-1}(1-\alpha)\}, \tag{3.3.4}$$

where $\chi^2_{q-1}(1-\alpha)$ is $(1-\alpha)th$ quantile of the χ^2 distribution with $q-1$ degrees of freedom. Let

$$P\Big[\max_{i,j} |y_i - y_j| < c(q, 1-\alpha) \Big| \mathbf{y} \sim N_q(\mathbf{0}, \mathbf{I}_q) \Big] = 1 - \alpha, \tag{3.3.5}$$

this constant $c(q, 1-\alpha)$ can be found in the Tables of the monograph Hollander and Wolfe (1999) or of Harter (1960). If

$$\min\{n_1, \ldots, n_q\} \to \infty, \quad \frac{n_i}{n} \to p_i > 0, \ \ i = 1, \ldots, q, \tag{3.3.6}$$

$$\Delta(R, n_1, \ldots, n_q) = \max_{i,j} \frac{\Big| \frac{S_{[i]}(R)}{n_i} - \frac{S_{[j]}(R)}{n_j} \Big|}{\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sqrt{\frac{24}{n(n+1)}} \tag{3.3.7}$$

and the random vector (3.3.2) is uniformly distributed, then the convergence

$$P\Big(\Delta(R, n_1, \ldots, n_q) > c(q, 1-\alpha) \Big) \to \gamma \le \alpha \tag{3.3.8}$$

holds. Here γ depends on the limiting values $p_1, \ldots, p_q$ only and the inequality in (3.3.8) holds with the equality sign if $p_1 = \ldots = p_q = \frac{1}{q}$. This can be proved by means of the results of Critchlow and Fligner (1991), pp. 130–131, the proof can be found on pp. 731 – 732 of Rublík (2005).

Theorem 3.3.1. *Suppose that* $\breve{\mathbf{Z}}$ *denotes transform of* $\mathbf{Z}$ *computed by means of (2.4.5). By means of the dissimilarity* t *compute (cf. (2.3.25)), (2.3.23))*

$$\overline{R}^{(NNst)} = R^{(NNst)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n, \mathbf{M}), \quad \overline{R}^{(NNt)} = R^{(NNt)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n, \mathbf{M}),$$

where $\mathbf{M} = \mathbf{0}$ *if all vectors* $\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n$ *are different from* $\mathbf{0}$*, and* $\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n} \breve{\mathbf{Z}}_i$ *otherwise. Let* $\overline{R}^{(NNst)(1)} < \ldots < \overline{R}^{(NNst)(n)}$ *denote the ordering of the first* n *coordinates of* $\overline{R}^{(NNst)}$ *according to their magnitude and similarly,* $\overline{R}^{(NNt)(1)} < \ldots < \overline{R}^{(NNt)(n)}$ *denote the ordering of the first* n *coordinates of* $\overline{R}^{(NNt)}$*. Further, let*

$$c_{NNst} = \sum_{j=1}^{n} \Big(\overline{R}^{(NNst)(j)} - \frac{n+1}{2} \Big) \Big(\overline{R}^{(NNst)(n+1-j)} - \frac{n+1}{2} \Big), \tag{3.3.9}$$

$$c_{NNt} = \sum_{j=1}^{n} \Big(\overline{R}^{(NNt)(j)} - \frac{n+1}{2} \Big) \Big(\overline{R}^{(NNt)(n+1-j)} - \frac{n+1}{2} \Big), \tag{3.3.10}$$

$$\tilde{R}^{(t)} = (\tilde{R}_1^{(t)}, \ldots, \tilde{R}_n^{(t)}) = \begin{cases} \breve{R}^{(NNt)} & |c_{NNt}| \ge |c_{NNst}|, \\ \breve{R}^{(NNst)} & \textit{otherwise}, \end{cases} \tag{3.3.11}$$

where $\breve{R}^{(NNt)} = R^{(NNt)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n)$*,* $R^{(NNst)} = R^{(NNst)}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n)$ *are the ranks computed as in (2.3.23) and (2.3.25).*

(I) Use the notation (3.3.2), (3.3.3) and put

$$\tilde{Q}_{n_1,\ldots,n_q,t} = Q_{n_1,\ldots,n_q}(\tilde{R}^{(t)}). \tag{3.3.12}$$

Since the assumptions imposed at the beginning of Section 3 hold, this quadratic test statistic of the Kruskal-Wallis type is uniquely defined with probability 1.

(II) Let the hypothesis (3.2) hold. Then the vector of ranks (3.3.11) is uniformly distributed over the set of all permutations $\mathcal{R}^{(n)}$ *and the test statistic (3.3.12) is as distributed as the Kruskal-Wallis q-sample rank test statistic in the case of the validity of the null hypothesis and sampling from one-dimensional continuous distribution. Hence if (3.3.6) is fulfilled then* $\tilde{Q}_{n_1,\ldots,n_q,t}$ *converges in distribution to the chi-square distribution with* $q-1$ *degrees of freedom and the convergence (3.3.8) holds with* $\Delta(\tilde{R}^{(t)}, n_1, \ldots, n_q)$.

Proof. The existence of the statistic (3.3.12) follows from Theorem 2.4.1(I). Similarly as in the proof of Theorem 3.1.2, the validity of Theorem 2.4.1 implies that the vector of ranks (3.3.11) is uniformly distributed over the set of permutations $\mathcal{R}^{(n)}$, and the rest of the proof follows from the previously mentioned results. □

In accordance with the previous Theorem the hypothesis (3.2) can be tested by means of the rejection region $Re(Q_{n_1,\ldots,n_q}(\tilde{R}^{(t)}))$ from (3.3.4), the simulation estimates of the probability of this region will be denoted by $\hat{P}_{\tilde{Q}_t}$. After the rejection of this null hypothesis the means $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_j$ are declared to be different if for (3.3.7) the inequality $\Delta(\tilde{R}^{(t)}, n_1, \ldots, n_q) > c(q, 1-\alpha)$ holds. Let *gd* denote the good decision that by this rule at least one pair of different location parameters was detected as different, and *wd* the wrong decision that at least one pair of identical location parameters was wrongly declared as different. The quality of this multiple comparisons rule can be described by the probabilities $P(gd|rej(\tilde{Q}_t))$ and $P(wd|rej(\tilde{Q}_t))$, $rej(\tilde{Q}_t) = Re(Q_{n_1,\ldots,n_q}(\tilde{R}^{(t)}))$, their simulation estimates are denoted by $\hat{P}_{\tilde{Q}_t}(gd)$ and $\hat{P}_{\tilde{Q}_t}(wd)$, respectively.

In the same way as in the previous Sections, the ranking of the observations can be carried out also means of the ordering based on the multiplication with a random vector. As details are obvious, we include only a brief description. Let $\breve{\mathbf{Z}}$ denote the transformation (2.4.5) of the pooled random sample $\mathbf{Z}$. Similarly as in Theorem 3.1.1 let $\mathbf{c} = \mathbf{c}_{FPP}(\breve{\mathbf{Z}}_1, \ldots, \breve{\mathbf{Z}}_n)$ be the projection pursuit vector (4.8) and $R^{(FPP)} = (R_1^{(FPP)}, \ldots, R_n^{(FPP)})$ denote the ranks of the numbers $\mathbf{c}^\top \breve{\mathbf{Z}}_1, \ldots, \mathbf{c}^\top \breve{\mathbf{Z}}_n$ computed by means of the ordering (2.2.1). Then $\hat{P}_{Q_{FPP}}$ will denote the simulation estimate of $P[Re(Q_{n_1,\ldots,n_q}(R^{(FPP)}))]$ of the rejection the null hypothesis (3.2) by means of the statistic $Q_{n_1,\ldots,n_q}(R^{(FPP)}))$, the quality of multiple comparisons rule based on the ranks $R^{(FPP)}$ will be described by the simulation estimates $\hat{P}_{\tilde{Q}_{FPP}}(gd)$ and $\hat{P}_{\tilde{Q}_{FPP}}(wd)$.

Finally, let $\tilde{\mathbf{c}} = \tilde{\mathbf{c}}(\mathbf{Z})$ be a non-zero characteristic vector corresponding to the largest characteristic root of the sample covariance matrix $\frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^\top$ of the pooled random sample (3.3.1), and $\mathbf{c}$ be the vector (2.3.4). Compute the scalar products $\mathbf{c}^\top \mathbf{Z}_1, \ldots, \mathbf{c}^\top \mathbf{Z}_n$ and use the ordering (2.2.1) to compute the ranks $R^{(pc)} = (R_1^{(pc)}, \ldots, R_n^{(pc)})$ of the pooled random sample $\mathbf{Z}$. Then $\hat{P}_{Q^{(pc)}}$ denotes the simulation estimate of the probability of rejection $P[Re(Q_{n_1,\ldots,n_q}(R^{(pc)}))]$ and $\hat{P}_{Q^{(pc)}}(gd)$, $\hat{P}_{Q^{(pc)}}(wd)$ denote simulation estimates of the probability of good decision and wrong decision, respectively.

In the following simulations the sample sizes $n_1 = 10$, $n_2 = 10$, $n_3 = 15$. Similarly as in the previous Section, the simulation results yielding (under the particular alternative) the largest power will be typed in bold, methods yielding the difference of the power not exceeding 0.06 will be regarded as equivalent.

Table 3.3.1. Sampling from $N_{1000}(\boldsymbol{\mu}_i, \mathbf{V})$ distribution, i=1,2,3, N=1000 trials.

	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1},\ i = 1,2,3$	$\boldsymbol{\mu}_i = c_i\,\mathbf{1}'_{1000},\ i = 1,2,3,$ $c_1 = -0.23, c_2 = 0, c_3 = 0.23$	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1}, i = 1,2,$ $\boldsymbol{\mu}_3 = 0.5\,\mathbf{1}'_{1000}$
$\hat{P}_{\tilde{Q}_{t_2}}$	0.048	**0.531**	**0.789**
$\hat{P}_{\tilde{Q}_{t_2}}(gd)$	0	0.951	0.949
$\hat{P}_{\tilde{Q}_{t_2}}(wd)$	0.102	0	0
$\hat{P}_{Q_{FPP}}$	0.039	0.414	0.634
$\hat{P}_{\tilde{Q}_{FPP}}(gd)$	0	0.937	0.953
$\hat{P}_{\tilde{Q}_{FPP}}(wd)$	0.128	0	0
$\hat{P}_{Q^{(pc)}}$	0.047	**0.537**	**0.794**
$\hat{P}_{Q^{(pc)}}(gd)$	0	0.957	0.945
$\hat{P}_{Q^{(pc)}}(wd)$	0.106	0	0
$\hat{P}_{\tilde{Q}_{t_1}}$	0.049	**0.529**	**0.789**
$\hat{P}_{\tilde{Q}_{t_1}}(gd)$	0	0.958	0.951
$\hat{P}_{\tilde{Q}_{t_1}}(wd)$	0.102	0	0

Table 3.3.2. Sampling from Cauchy $C_{1000}(\boldsymbol{\mu}_i, \mathbf{V})$ distribution, i=1,2,3, N=1000 trials.

	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1},\ i = 1,2,3$	$\boldsymbol{\mu}_i = c_i\,\mathbf{1}'_{1000},\ i = 1,2,3,$ $c_1 = -16, c_2 = 0, c_3 = 16$	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1}, i = 1,2,$ $\boldsymbol{\mu}_3 = 23\,\mathbf{1}'_{1000}$
$\hat{P}_{\tilde{Q}_{t_2}}$	0.042	**0.949**	**0.895**
$\hat{P}_{\tilde{Q}_{t_2}}(gd)$	0	0.992	0.972
$\hat{P}_{\tilde{Q}_{t_2}}(wd)$	0.095	0	0
$\hat{P}_{Q_{FPP}}$	0.034	0.593	0.483
$\hat{P}_{\tilde{Q}_{FPP}}(gd)$	0	0.948	0.917
$\hat{P}_{\tilde{Q}_{FPP}}(wd)$	0.177	0	0
$\hat{P}_{Q^{(pc)}}$	0.041	0.595	0.492
$\hat{P}_{Q^{(pc)}}(gd)$	0	0.950	0.902
$\hat{P}_{Q^{(pc)}}(wd)$	0.073	0	0
$\hat{P}_{\tilde{Q}_{t_1}}$	0.048	**0.926**	**0.872**
$\hat{P}_{\tilde{Q}_{t_1}}(gd)$	0	0.990	0.966
$\hat{P}_{\tilde{Q}_{t_1}}(wd)$	0.042	0	0

Table 3.3.3. Sampling from chi-square $\chi^2_{1000,5}(\boldsymbol{\mu}_i, \mathbf{V})$ distribution, i=1,2,3, N=1000 trials.

	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1},\ i = 1,2,3$	$\boldsymbol{\mu}_i = c_i\,\mathbf{1}'_{1000},\ i = 1,2,3,$ $c_1 = -0.9, c_2 = 0, c_3 = 0.9$	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1}, i = 1,2,$ $\boldsymbol{\mu}_3 = 1.5\,\mathbf{1}'_{1000}$
$\hat{P}_{\tilde{Q}_{t_2}}$	0.051	**0.726**	**0.700**
$\hat{P}_{\tilde{Q}_{t_2}}(gd)$	0	0.961	0.936
$\hat{P}_{\tilde{Q}_{t_2}}(wd)$	0.059	0	0
$\hat{P}_{Q_{FPP}}$	0.047	0.651	0.630
$\hat{P}_{\tilde{Q}_{FPP}}(gd)$	0	0.951	0.932
$\hat{P}_{\tilde{Q}_{FPP}}(wd)$	0.128	0	0
$\hat{P}_{Q^{(pc)}}$	0.050	**0.728**	**0.712**
$\hat{P}_{Q^{(pc)}}(gd)$	0	0.975	0.938
$\hat{P}_{Q^{(pc)}}(wd)$	0.120	0	0
$\hat{P}_{\tilde{Q}_{t_1}}$	0.052	**0.721**	**0.699**
$\hat{P}_{\tilde{Q}_{t_1}}(gd)$	0	0.968	0.936
$\hat{P}_{\tilde{Q}_{t_1}}(wd)$	0.039	0	0

Table 3.3.4. Sampling from the generalized Pareto $G_{1000}(\boldsymbol{\mu}_i, \mathbf{V}, 1, 1, 1)$ distribution, $i = 1, 2, 3$, N=1000 trials.

	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1}$, $i = 1, 2, 3$	$\boldsymbol{\mu}_i = c_i\,\mathbf{1}'_{1000}$, $i = 1, 2, 3$, $c_1 = -23, c_2 = 0, c_3 = 23$	$\boldsymbol{\mu}_i = \mathbf{0}_{d\times 1}, i = 1, 2$, $\boldsymbol{\mu}_3 = 40\,\mathbf{1}'_{1000}$
$\hat{P}_{\tilde{Q}_{t_2}}$	0.047	**0.608**	**0.693**
$\hat{P}_{\tilde{Q}_{t_2}}(gd)$	0	0.961	0.951
$\hat{P}_{\tilde{Q}_{t_2}}(wd)$	0.064	0	0
$\hat{P}_{Q_{FPP}}$	0.038	0.530	0.555
$\hat{P}_{\tilde{Q}_{FPP}}(gd)$	0	0.957	0.941
$\hat{P}_{\tilde{Q}_{FPP}}(wd)$	0.184	0	0.002
$\hat{P}_{Q^{(pc)}}$	0.038	0.557	0.581
$\hat{P}_{Q^{(pc)}}(gd)$	0	0.959	0.945
$\hat{P}_{Q^{(pc)}}(wd)$	0.184	0	0
$\hat{P}_{\tilde{Q}_{t_1}}$	0.047	**0.640**	**0.717**
$\hat{P}_{\tilde{Q}_{t_1}}(gd)$	0	0.956	0.957
$\hat{P}_{\tilde{Q}_{t_1}}(wd)$	0.064	0	0.000

These simulation results suggest that for the considered distributions, dimension of observations and sample sizes, the performance of the statistics $Q_{n_1,\ldots,n_q,t_2}$, $Q_{n_1,\ldots,n_q,t_1}$ is equivalent. An equivalent power for sampling from the normal or chi-square distribution is achieved also by the statistic $Q_{n_1,\ldots,n_q}(R^{(pc)})$ based on the ranks of the first principal component, but for sampling from Cauchy distribution or from generalized Pareto distribution the statistic $Q_{n_1,\ldots,n_q}(R^{(pc)})$ is outperformed both by $Q_{n_1,\ldots,n_q,t_2}$ and $Q_{n_1,\ldots,n_q,t_1}$.

Obviously, the description of performance of tests in this multisample problem will require further extensive simulations.

4. Conclusion

As the multisample case was in the previous Subsection handled only tentatively for several alternatives, we pay attention to the two-sample testing of the location hypothesis. As is explained at the end of Subsection 3.2, two amongst the considered tests appear to yield for the considered moderate sample sizes the best results: the test of Feng, Zou and Wang (briefly FZW), and also the test based on the Wilcoxon test statistic, the ranks (3.1.9) and the dissimilarity t_2 from (2.5.4).

When comparing the power of these tests in the framework of the completely unknown continuous distribution, the performance of the FZW test and the test based on the dissimilarity t_2 is approximately the same. As to the case of the special type of distributions, as an advantage of the FZW test one may regard the fact that it yields better results when sampling is drawn from Gaussian distributions. As can be seen from the simulation results presented in Subsection 3.2, the test based on t_2 is better for sampling from the Cauchy distribution and from the generalized Pareto distribution, the FZW test is better for sampling from the normal distribution and from the chi-square distribution.

A slight disadvantage of FZW test is that it uses an iterative computational procedure for which the exact proof of convergence is unknown, but this may be handled in the way explained in the description of the simulations of its power (cf. the text following the Table 3.2.26). An important feature of the t_2 based test is that it is based on the precise null distribution and yields precise chosen significance level.

As is explained at the end of Subsection 3.2, when one wants to have a safeguard against the case when the scale parameter of some coordinate of observations strikingly exceeds the scale parameters of the other coordinates, then for high dimensions (like d=1000) the use of the test based on the Friedman projection pursuit vector is recommendable.

As far as some extensions of the results of the paper are concerned, one could use the fact that similarly as in the construction of the sample principal components, the construction of the extreme values of the dissimilarity based ranks proposed in this paper begins with the use of the first characteristic vector of the sample covariance matrix. As has already been in the paper said, it is not clear whether the first principal component should separate the groups best and one could also consider a use of more than one component for the construction of the multivariate ranks, which is a possible matter of further research.

Appendix: Computation of the projection pursuit vector

For the sake of completeness, we include the description of the algorithm for computation of the Friedman projection pursuit vector, used in the simulations mentioned in the previous text. It is based on the formulas of Friedman (1987). Even though this description uses Matlab commands, it is not an algorithm ready for use, and for the space reasons it provides an outline of computation procedure only. The computation similarly as the approach of Friedman and Tukey (1974 ser. C) uses a transformation of the unit sphere to an Euclidean space, which enables to proceed without constraints.

The computation is carried out in three parts. In Part 1 the procedure (33) of Friedman (1987) is used to obtain a preliminary initial value $\mathbf{a}$ of the projection pursuit vector. In Part 2 the initial value $\boldsymbol{\varphi}^{(0)}$ of the spherical coordinates, used in the next iterations, is computed (i.e, if all coordinates of $\mathbf{a}$ are different from zero, then $\boldsymbol{\varphi}^{(0)}$ are its spherical coordinates, otherwise $\boldsymbol{\varphi}^{(0)}$ are spherical coordinates of a suitable approximation of $\mathbf{a}$). In Part 3 an iterative computation based on the $\boldsymbol{\varphi}^{(0)}$ is carried out and the resulting value is by means of (4.8) transformed back into the unit sphere.

As the procedure is based on the use of the gradient method and employs the multivariate spherical coordinates, let us recall their definition. Let $T_d = T_d(\boldsymbol{\varphi}) : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the mapping with the coordinates $cos(\varphi_1)$, $sin(\varphi_1)cos(\varphi_2)$, ..., $\Big(\prod_{j=1}^{d-2} sin(\varphi_j)\Big)cos(\varphi_{d-1})$ and the last coordinate is $\Big(sin(\varphi_1)sin(\varphi_2)\dots sin(\varphi_{d-2})sin(\varphi_{d-1})\Big)$. Then $\|T_d(\boldsymbol{\varphi}\| = 1$ and for every vector $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ with $\|\mathbf{x}\| = 1$ and $\prod_{i=1}^{d} x_i \neq 0$ there exists a unique vector $\boldsymbol{\varphi} \in (0, 2\pi) \times (0, \pi)^{d-2}$ such that $\mathbf{x} = T_d(\boldsymbol{\varphi})$. This vector is described by the formulas

$$\varphi_{d-1} = arccotg\Big(\frac{x_{d-1}}{x_d}\Big), \quad \varphi_j = arccotg\Big(\frac{x_j}{x_{j+1}} cos(\varphi_{j+1})\Big), \; j = d-2, \dots, 2 \tag{4.1}$$

$$\left.\begin{array}{l} x_1 > 0,\, x_2 > 0,\, cos(\varphi_2) > 0 \\ x_1 > 0,\, x_2 < 0,\, cos(\varphi_2) < 0 \\ x_1 < 0,\, x_2 > 0,\, cos(\varphi_2) > 0 \\ x_1 < 0,\, x_2 < 0,\, cos(\varphi_2) < 0 \end{array}\right\} \varphi_1 = arccotg\Big(\frac{x_1}{x_2} cos(\varphi_2)\Big), \tag{4.2}$$

$$\left.\begin{array}{l} x_1 > 0, x_2 > 0, cos(\varphi_2) < 0 \\ x_1 > 0, x_2 < 0, cos(\varphi_2) > 0 \\ x_1 < 0, x_2 < 0, cos(\varphi_2) > 0 \\ x_1 < 0, x_2 > 0, cos(\varphi_2) < 0 \end{array}\right\} \varphi_1 = arccotg\Big(\frac{x_1}{x_2} cos(\varphi_2)\Big) + \pi. \tag{4.3}$$

where $arccotg : (-\infty, +\infty) \to (0, \pi)$. For this vector $\boldsymbol{\varphi}$ we shall use in the next text the notation $\boldsymbol{\varphi} = sph(\mathbf{x})$.

Let $\mathbf{z}_1, \dots, \mathbf{z}_n$, $\mathbf{a}$ be vectors from $\mathbb{R}^d$ and $\|\mathbf{a}\| = 1$. In accordance with the formula (10) and p. 258 of Friedman (1987) let

$$\hat{I}(\mathbf{a}) = \frac{1}{2} \sum_{j=1}^{6} (2j+1) \left[\frac{1}{n} \sum_{i=1}^{n} P_j(2\Phi(\mathbf{a}^\top \mathbf{z}_i) - 1)\right]^2 \tag{4.4}$$

denote the projection index of order 6, where $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2 - 1)$,...,$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$ are the first 6 Legendre polynomials and Φ denotes the distribution function of the $N(0,1)$ distribution. The aim is to maximize $\hat{I}(\mathbf{a})$.

Part 1. In this part the projection index is preliminarily maximized by means of the coordinate axes. As before, let $\mathbf{e}_i \in \mathbb{R}^d$ denote the vector whose j-th coordinate is 1 if $j = i$ and 0 otherwise. In accordance with (33) of Friedman (1987) let

$$j^* = \min\Big\{j; \hat{I}(\mathbf{e}_j) = \max_{1\le\ell\le d} \hat{I}(\mathbf{e}_\ell)\Big\},\ \hat{I}_0 = \hat{I}(\mathbf{e}_{j^*}),\quad \mathbf{a}^* = \mathbf{e}_{j^*}.$$

$$f(j^*) = \hat{I}(\mathbf{e}_{j^*}),\ s(j^*) = 0,\ k = 1,\ \mathbf{A}(:,k) = \mathbf{a}^*.$$

For $i = 1,\dots,q$, $i \ne j^*$ let

$$f_+ = I\Big(\tfrac{\mathbf{a}^*+\mathbf{e}_i}{\sqrt{2}}\Big),\quad f_- = I\Big(\tfrac{\mathbf{a}^*-\mathbf{e}_i}{\sqrt{2}}\Big),$$
$$if\ f_+ > f_-\ then\ f(i) = f_+,\ s(i) = 1,\ else\ f(i) = f_-,\ s(i) = -1,$$
$$f_m = \max_\ell f(\ell).$$

Choose the inaccuracy $\epsilon > 0$ (in the simulations of the paper $\epsilon = 0.01$) and put

$$Df = \chi_{\langle\epsilon,+\infty)}(f_m - \hat{I}_0),\ difference(1) = 1.$$

where $\chi_{\langle\epsilon,+\infty)}(y)$ equals 1 if $y \ge \epsilon$, and 0 otherwise. The next procedure is constructed in such a way that its length does not exceed d.

while $Df > 0$
  $f_0 = \hat{I}(\mathbf{a}^*),\ f_{mold} = f_m,$
    for $i = 1 \dots q$
      $f_+ = \hat{I}\Big(\dfrac{\mathbf{a}^* + \mathbf{e}_i}{\|\mathbf{a}^* + \mathbf{e}_i\|}\Big),\ f_- = \hat{I}\Big(\dfrac{\mathbf{a}^* - \mathbf{e}_i}{\|\mathbf{a}^* - \mathbf{e}_i\|}\Big),$
      if $f_+ > f_-$ *then* $f(i) = f_+,\ s(i) = 1$ *else* $f(i) = f_-,\ s(i) = -1$
    end;
$f_m = \max\limits_\ell f(\ell),\ i_1 = \min\{i; f(i) = f_m\},$
$k := k + 1,\ \mathbf{a}^* = \dfrac{\mathbf{a}^* + s(i_1)\mathbf{e}_{i_1}}{\|\mathbf{a}^* + s(i_1)\mathbf{e}_{i_1}\|},\ \mathbf{A}(:,k) = \mathbf{a}^*,\ difference(k) = sign(f_m - f_{mold})$
$Df = \chi_{\langle\epsilon,+\infty)}(f_m - f_{mold})\chi_{(-\infty,d\rangle}(k)$
end

where $\chi_{(-\infty,d\rangle}(k)$ equals 1 if $k \le d$ and 0 otherwise. Taking into account the possibility of the decrease in the last iteration we put $\mathbf{a} = \mathbf{A}(:,k-1)$ if $difference(k) = -1$, otherwise $\mathbf{a} = \mathbf{A}(:,k)$.

Part 2. In Part 3 we shall use the gradient method implemented by means of the spherical coordinates. As they are uniquely defined only for vectors with nonzero coordinates, in this Part we shall find spherical coordinates $\boldsymbol{\varphi}^{(0)}$ which create a vector whose projection index equals (up to a chosen inaccuracy) the number $\hat{I}(\mathbf{a})$, where $\mathbf{a}$ is the vector obtained in the Part 1 of this procedure.

If $\prod_{i=1}^q \mathbf{a}(i) \ne 0$, put

$$\boldsymbol{\varphi}^{(0)} = sph(\mathbf{a}).$$

Let $\mathbf{a}(i) = 0$ for some i (i.e., the spherical coordinates are not defined). We are going to find non-zero angles $\boldsymbol{\varphi}^{(0)}$ such that $\hat{I}(T_d(\boldsymbol{\varphi}^{(0)})) \ge \hat{I}(\mathbf{a}) - 0.01$. Put

$$k = 10,\ \mathbf{t} = \Big(\delta(\mathbf{a}(1),0),\dots,\delta(\mathbf{a}(q),0)\Big)^\top$$

where $\delta(\alpha,\beta)=1$ if $\alpha=\beta$ and $\delta(\alpha,\beta)=0$ otherwise, denotes the Kronecker delta. We use the algorithm

$$\begin{aligned}
&\mathbf{a}^*=\Big(1-\tfrac{1}{4^k}\Big)\mathbf{a}+\tfrac{1}{4^k}\mathbf{t},\quad \tilde{\mathbf{a}}=\tfrac{\mathbf{a}^*}{\|\mathbf{a}^*\|},\\
\texttt{while}\quad &\hat{I}(\mathbf{a}^*)<\big(\hat{I}(\mathbf{a})-0.01\big)\&(k<25)\\
&k=k+1,\ \mathbf{a}^*=\Big(1-\tfrac{1}{4^k}\Big)\mathbf{a}+\tfrac{1}{4^k}\mathbf{t},\quad \tilde{\mathbf{a}}=\tfrac{\mathbf{a}^*}{\|\mathbf{a}^*\|},\\
\texttt{end}&
\end{aligned}$$

i.e., the search for $\tilde{\mathbf{a}}$ will stop not later than after 25 steps, which will be sufficient in the majority of cases. By means of this vector compute

$$\boldsymbol{\varphi}^{(0)}=sph(\tilde{\mathbf{a}}).$$

Part 3. We shall use in this Part the notation

$$\hat{I}_s(\boldsymbol{\varphi})=\hat{I}\Big(T_d(\boldsymbol{\varphi})\Big),\quad \nabla\hat{I}_s(\boldsymbol{\varphi})=\begin{pmatrix}\frac{\partial\hat{I}_s(\boldsymbol{\varphi})}{\partial\varphi_1}\\ \vdots\\ \frac{\partial\hat{I}_s(\boldsymbol{\varphi})}{\partial\varphi_{d-1}}\end{pmatrix}.$$

Compute $M=\max\{\Big|\frac{\partial\hat{I}_s(\boldsymbol{\varphi})}{\partial\varphi_i}\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}^{(0)}};i=1,\dots,d-1\}$. If $M=0$ then $\boldsymbol{\varphi}^{(0)}$ is a stationary point of $\hat{I}$. Therefore in this case put

$$\hat{\boldsymbol{\varphi}}=\boldsymbol{\varphi}^{(0)},\tag{4.5}$$

and stop the iteration procedure.

Let $M>0$. Define the column vector by the formula

$$\mathbf{t}=\Big(sign\Big(\frac{\partial\hat{I}_s(\boldsymbol{\varphi})}{\varphi_i}|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}^{(0)}}\Big);i=1,\dots,d-1\Big).$$

It is known that for all $\alpha>0$ sufficiently small and $\tilde{\boldsymbol{\varphi}}=\boldsymbol{\varphi}^{(0)}+\alpha\mathbf{t}$ the inequality $\hat{I}_s(\tilde{\boldsymbol{\varphi}})>\hat{I}_s(\boldsymbol{\varphi}^0)$ holds. We shall find such a majorizing point by means of the following iteration computation.

$$\begin{aligned}
&k=2;\ \ firstAngles=\boldsymbol{\varphi}^{(0)}+\tfrac{1}{10^k}\mathbf{t};\ \ Ik=\hat{I}_s(firstAngles);\\
\texttt{while}\quad &Ik<(\hat{I}_s(\boldsymbol{\varphi}^0)-0.001)\&(k<17);\\
&k:=k+1;firstAnglesk=\boldsymbol{\varphi}^{(0)}+\tfrac{1}{10^k}\mathbf{t};\quad Ik=\hat{I}_s(firstAnglesk);\\
\texttt{end}&\\
\texttt{if}\quad &k==2\\
&\boldsymbol{\varphi}^{(1)}=firstAngles\\
\texttt{else}&\\
&\boldsymbol{\varphi}^{(1)}=firstAnglesk;\\
\texttt{end}&
\end{aligned}$$

i.e., the search for the angles $\boldsymbol{\varphi}^{(1)}=\tilde{\boldsymbol{\varphi}}$ will be stopped not later than after 17 steps, which will be sufficient in the majority of cases.

Suppose that $k\geq 1$ and vector $\boldsymbol{\varphi}^{(k)}$ is already defined. Then

$$if\ \ \nabla\hat{I}_s(\boldsymbol{\varphi}^{(k-1)})=\nabla\hat{I}_s(\boldsymbol{\varphi}^{(k)})\ put\ \hat{\boldsymbol{\varphi}}=\boldsymbol{\varphi}^{(k)}\ and\ stop\ the\ computation.\tag{4.6}$$

If $\nabla \hat{I}_s(\boldsymbol{\varphi}^{(k-1)}) \neq \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)})$, then define

$$\gamma_k = \frac{(\boldsymbol{\varphi}^{(k)} - \boldsymbol{\varphi}^{(k-1)})^\top}{\|\nabla \hat{I}_s(\boldsymbol{\varphi}^{(k-1)}) - \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)})\|^2} \Big[\nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)}) - \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k-1)})\Big],$$

$$\boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} + \gamma_k \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)}) \; if \; \hat{I}_s\Big(\boldsymbol{\varphi}^{(k)} + \gamma_k \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)})\Big) > \hat{I}_s(\boldsymbol{\varphi}^{(k)}),$$

$$otherwise \; \boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - \gamma_k \nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)}).$$

Employ these formulas and use the stopping rule

$$if \;\; \|\nabla \hat{I}_s(\boldsymbol{\varphi}^{(k)})\| \leq 0.01 \; or \; k \geq 37 \; put \; \hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}^{(k)} \; and \; stop \; the \; computation. \tag{4.7}$$

Finally, making use of (4.5)–(4.7) compute the Friedman project pursuit vector

$$\mathbf{c}_{FPP}(\mathbf{z}_1, \ldots, \mathbf{z}_n) = T_d(\hat{\boldsymbol{\varphi}}), \tag{4.8}$$

where the mapping T_d is defined at the beginning of this Section.

It is obvious that in the stopping rules mentioned in this Section the inaccuracies and the number of iterations may be chosen according to one's personal preferences. Although the convergence of this procedure is not guaranteed, this construction of computation yields the chosen upper bound for the number of iterations (in the present text, the bound is 37).

References

Barnett V (1976). "The Ordering of Multivariate Data." *Journal of Royal Statistical Society, A*, **139**(3), 318–355.

Biswas M, Ghosh AK (2014). "A Nonparametric Two-Sample Test Applicable to High Dimensional Data." *Journal of Multivariate Analysis*, **123**, 160–171.

Chen SX, Qin YL (2010). "A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing." *The Annals of Statistics*, **38**(2), 808–835.

Critchlow DE, Fligner MA (1991). "On Distribution-Free Multiple Comparisons in the One-Way Analysis of Variance." *Communications in Statistics-Theory and Methods*, **20**(1), 127–139.

Feng L, Sun F (2016). "Spatial-Sign Based High-Dimensional Location Test." *Electronic Journal of Statistics*, **10**(2), 2420–2434. URL DOI:10.1214/16-EJS1176.

Feng L, Zou C, Wang Z (2016). "Multivariate-Sign-Based High-Dimensional Tests for the Two-Sample Location Problem." *Journal of the American Statistical Association*, **111**(514), 721–735. URL DOI:10.1080/01621459.2015.1035380.

Fraiman R, Moreno L, Vallejo S (2017). "Some Hypothesis Test Based on Random Projection." *Computational Statistics*, **32**(3), 1165–1189. URL DOI10.1007/s00180-017-0732-4.

Friedman JH (1987). "Exploratory Projection Pursuit." *Journal of the American Statistical Association*, **82**(397), 249–266.

Friedman JH, Tukey JW (1974 ser. C). "A Projection Pursuit Algorithm for Exploratory Data Analysis." *IEEE Transactions on Computers*, **23**(9), 881–889.

Hájek J, Šidák Z, Sen PK (1999). *Theory of Rank Tests*. Academic Press, New York. ISBN 0-12-642350-4.

Harter HL (1960). "Tables of Range and Studentized Range." *The Annals of Mathematical Statistics*, **31**(4), 1122 – 1147.

Hollander M, Wolfe DA (1999). *Nonparametric Statistical Methods*. Wiley, New York. ISBN 0-471-19045-4.

Kruskal WH, Wallis WA (1952). "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association*, **47**(260), 583–621.

Nordhausen K, Oja H, Tyler DE (2006). "On the Efficiency of Invariant Multivariate Sign and Rank Tests." *Festschrift for Tarmo Pukkila on his 60th birthday*, p. 217.

Oja H (2010). *Multivariate Nonparametric Methods with R*. Springer, New York. DOI 10.1007/978-1-4419-0468-3.

Oja H, Randles RH (2004). "Multivariate Nonparametric Tests." *Statistical Science*, **19**(4), 598–605. URL DOI10.1214/088342304000000558.

Rousson V (2002). "On Distribution-Free Tests for the Multivariate Two-Sample Location-Scale Model." *Journal of Multivariate Analysis*, **80**, 43–57. URL doi:10.1006/jmva.2000.1981.

Rublík F (2005). "The Multisample Version of the Lepage test." *Kybernetika*, **41**(6), 713 – 733.

Wang L, Peng B, Li R (2015). "A High-Dimensional Nonparametric Multivariate Test for Mean Vector." *Journal of the American Statistical Association*, **110**(512), 1658– 1669. URL https://doi.org/10.1080/01621459.2014.988215.

Affiliation:

František Rublík
Institute of Measurement Science
Slovak Academy of Sciences
84104 Bratislava, Slovakia
E-mail: umerrubl@savba.sk
URL: http://www.um.sav.sk/en/department-03/frantisek-rublik.html