



Common multivariate estimators of location and scatter capture the symmetry of the underlying distribution

Jan Kalina^{a,b} 

^aThe Czech Academy of Sciences, Institute of Computer Science, Prague, Czech Republic; ^bThe Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czech Republic

ABSTRACT

The article discusses how various multivariate location and scatter estimators capture the symmetry of the underlying distribution. Very general sufficient conditions are formulated, which ensure various symmetry properties of functionals corresponding to location or scatter. Examples of robust multivariate estimators, which fulfill these conditions, are discussed in detail. The obtained symmetry of the estimators is applicable to hypothesis tests of symmetry of the underlying distribution of the multivariate data. For this task, we propose to perform permutation tests exploiting the nonparametric combination methodology. The performance of the newly proposed tests is illustrated on simulated as well as real data. The tests are suitable for small sample sizes and represent the first available symmetry tests suitable also for non-elliptical distributions and for more than just two variables.

ARTICLE HISTORY

Received 26 March 2018
Accepted 1 May 2019



KEYWORDS

Multivariate estimation; Symmetry test; Robust estimation; Scatter estimator; Axial symmetry

1. Introduction

Let us consider a random vector $\mathbf{X} \in \mathbb{R}^p$. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be p -dimensional i.i.d. random vectors following the probability distribution $\mathcal{L}(\mathbf{X})$. The aim is to estimate parameters $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Delta} \in \mathbb{R}^{p \times p}$, which are related to the location and scatter of \mathbf{X} . Special cases of $\boldsymbol{\Delta}$ include the covariance matrix $\boldsymbol{\Sigma}$ or the shape matrix defined as $\boldsymbol{\Sigma}/(\det(\boldsymbol{\Sigma}))^{1/p}$, if these matrices exist and where $\det(\boldsymbol{\Sigma})$ denotes the determinant of $\boldsymbol{\Sigma}$.

While numerous estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Delta}$ based on the random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ are available, this paper investigates symmetry properties of their corresponding functionals with an application to hypothesis testing about various forms of distributional symmetry. Several leading concepts of multivariate symmetry will be considered. Serfling (2006) or Oja (2010) overviewed standard multivariate symmetry concepts in increasing order of generality as spherical, elliptical, central, and angular symmetry. In addition, marginal symmetry is defined as the symmetry of all the marginal distributions around a point. Symmetry of functionals corresponding to multivariate estimators was thoroughly described by Tatsuoka and Tyler (2000). In addition, symmetry of the underlying distribution is an important assumption for various robust multivariate estimators.

CONTACT Jan Kalina  kalina@cs.cas.cz  The Czech Academy of Sciences, Institute of Computer Science, Prague, Czech Republic; The Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czech Republic.

Section 2 of this paper presents the main result devoted to functionals corresponding to important robust multivariate estimators. Under suitable assumptions, the functionals are shown to capture the symmetry of the underlying distribution. Based on this, we propose to apply the nonparametric combination methodology to individual permutation tests of various forms of symmetry for multivariate data. Various robust multivariate estimators are presented in Sec. 3, which fulfill the assumptions of our main result. Numerical simulations are presented in Sec. 4, a real data example in Sec. 5, and conclusions in Sec. 6.

2. Symmetry of multivariate estimators

2.1. Notation

We introduce the notation $\text{PSD}(p)$ and $\text{PD}(p)$ for the set of all positive semidefinite symmetric and positive definite symmetric matrices of size $p \times p$, respectively, and consider two sets

$$\Theta = \{(\mathbf{m}, \mathbb{A}); \mathbf{m} \in \mathbb{R}^p, \mathbb{A} \in \text{PSD}(p)\} \quad \text{and} \quad \Psi = \{(\mathbf{m}, \mathbb{A}); \mathbf{m} \in \mathbb{R}^p, \mathbb{A} \in \text{PD}(p)\}.$$

Let us consider a general optimization problem well-defining an estimator $\mathbf{m}(\mathbf{X})$ of $\boldsymbol{\mu}$ and $\mathbb{A}(\mathbf{X})$ related to $\mathbf{\Delta}$ in the form

$$\min_{(\mathbf{m}, \mathbb{A}) \in \Theta} \mathbb{E} f_0(\mathbf{X}, \mathbf{m}, \mathbb{A}) \quad (2.1)$$

subject to

$$\begin{aligned} \mathbb{E} f_u(\mathbf{X}, \mathbf{m}, \mathbb{A}) &\leq 0, & u = 1, \dots, U, \\ \mathbb{E} h_v(\mathbf{X}, \mathbf{m}, \mathbb{A}) &= 0, & v = 1, \dots, V, \end{aligned}$$

where given functions f_0, f_1, \dots, f_U and h_1, \dots, h_V are considered on the domain

$$\left(\bigcap_{u=0}^U \text{dom } f_u \right) \cap \left(\bigcap_{v=1}^V \text{dom } h_v \right). \quad (2.2)$$

Prominent examples of $f_0(\mathbf{X}, \mathbf{m}, \mathbb{A})$ include functions of $(\mathbf{X} - \mathbf{m})^T \mathbb{A}^{-1} (\mathbf{X} - \mathbf{m})$ or functions of $\det(\mathbb{A})$.

2.2. Main result

Population counterparts of $\mathbf{m}(\mathbf{X})$ and $\mathbb{A}(\mathbf{X})$ obtained as functional solutions of (2.1) will be denoted as $\mathbf{m}_0(\mathbf{X})$ and $\mathbb{A}_0(\mathbf{X})$. The following set of assumptions will be considered.

Assumptions A.

1. $\mathbf{m}_0(\mathbf{X})$ is shift equivariant and rotation invariant;
2. $\mathbb{A}_0(\mathbf{X})$ is shift invariant and rotation equivariant;
- 3.

$$\begin{aligned} f_u(\mathbb{O}(\mathbf{X} - \mathbf{c}), \mathbb{O}(\mathbf{m} - \mathbf{c}), \mathbb{O}^T \mathbb{A} \mathbb{O}) &= f_u(\mathbf{X} - \mathbf{c}, \mathbf{m} - \mathbf{c}, \mathbb{A}), \quad u = 0, \dots, U, \\ h_v(\mathbb{O}(\mathbf{X} - \mathbf{c}), \mathbb{O}(\mathbf{m} - \mathbf{c}), \mathbb{O}^T \mathbb{A} \mathbb{O}) &= h_v(\mathbf{X} - \mathbf{c}, \mathbf{m} - \mathbf{c}, \mathbb{A}), \quad v = 1, \dots, V, \end{aligned}$$

for all orthonormal matrices \mathbb{O} and all vectors \mathbf{c} admissible with respect to (2.2);

4. The functionals \mathbf{m}_0 and \mathbb{A}_0 are uniquely defined (for the population case);
5. $\mathcal{L}(\mathbf{X}-\mathbf{c}) = \mathcal{L}(\mathbb{O}(\mathbf{X}-\mathbf{c}))$ for each shift vector \mathbf{c} and orthonormal matrix \mathbb{O} .

The following theorem ensures symmetry properties for functionals satisfying Assumptions A, particularly their ability to capture the symmetry of the underlying distribution.

Theorem 1. *We consider the estimators $\mathbf{m}(\mathbf{X})$ and $\mathbb{A}(\mathbf{X})$ defined by (2.1). If Assumptions A are fulfilled, then necessarily*

$$\mathbb{O}(\mathbf{m}-\mathbf{c}) = \mathbf{m}-\mathbf{c} \quad \text{and} \quad \mathbb{O}^T \mathbb{A}_0 \mathbb{O} = \mathbb{A}_0$$

for each shift vector \mathbf{c} and orthonormal matrix \mathbb{O} . Let us further denote $\mathbf{m}_0 = (m_1, \dots, m_p)^T$ and $\mathbb{A}_0 = (a_{ij})_{i,j=1}^p$. Let us assume $\mathcal{L}(\mathbf{X}) = \mathcal{L}(\mathbb{K}\mathbf{X})$ for a sign-change matrix $\mathbb{K} = \mathbb{K}^T = \mathbb{K}^{-1} = \text{diag}(k_1, \dots, k_p)$ with diagonal elements ± 1 . If $k_i = -1$ for any $i = 1, \dots, p$, then necessarily $m_i = 0$ for such i . If additionally $k_i k_j = -1$ for any pair $i, j = 1, \dots, p$, then $a_{ij} = 0$ for such i and j . Consequently,

- (I) if $\mathcal{L}(\mathbf{X})$ is centrally or marginally symmetric, then \mathbf{m}_0 coincides with the center of symmetry;
- (II) if $\mathcal{L}(\mathbf{X})$ is symmetric around an affine subspace (i.e. axis or hyperplane), then \mathbf{m}_0 lies on that affine subspace;
- (III) if $\mathcal{L}(\mathbf{X})$ is symmetric around the first coordinate axis, then $a_{ij} = 0$ for all $j = 2, \dots, p$;
- (IV) if $\mathcal{L}(\mathbf{X})$ is symmetric around linear subspace consisting of all the points with the first k coordinates zero, then $a_{ij} = 0$ for any $i \in 1, \dots, k$ and $j = k + 1, \dots, p$.

The proof is straightforward. It is worth noting that no restrictive assumptions are assumed, i.e. the result does not require continuity, differentiability, convexity, monotonicity, boundedness etc. The properties of the estimators may be preserved if they are computed recursively, i.e., the solution of one such optimization problem is used as an input to another problem of an analogous form.

2.3. Hypothesis testing

The statement of [Theorem 1](#) allows to construct hypothesis tests of various forms of symmetry of $\mathcal{L}(\mathbf{X})$. A literature research shows that available symmetry tests are mainly devoted to bivariate symmetry (i.e. bivariate exchangeability); see Rao and Raghunath (2012) or Quessy (2016). Numerous tests are available for spherical symmetry (Baringhaus 1991). There are also depth-based symmetry tests, e.g. of central symmetry by Paindaveine and Van Bever (2013), of angular symmetry about a specified center by Rousseeuw and Struyf (2002), or the test about an unspecified center by Dutta, Ghosh, and Chaudhuri (2011). A test of rotation symmetry on a hypersphere was proposed by García-Portugués, Paindaveine, and Verdebout (2018). Nevertheless, there have been no tests of axial symmetry proposed for $p \geq 2$ so far.

We focus on permutation tests and their nonparametric combination without any assumptions on the probabilistic distribution of the data. Permutation tests are very

general, simple and powerful. We will also present examples of tests exploiting particular estimators of the covariance matrix of multivariate data.

As an example, let us consider a test of H_0 that \mathbb{A} is a diagonal matrix against H_1 that H_0 is not true. We propose to use permutation tests for each individual a_{ij} for $i < j$ and to combine the individual tests by means of one of nonparametric approaches due to Fisher, Liptak or Tippett, which are standard tools of nonparametric combination methodology (see p. 147 of Pesarin and Salmaso (2010) or p. 4 of Bonnini et al. (2014)).

3. Examples: robust multivariate estimators

This section presents such important (robust) multivariate estimators of location and scatter, for which the corresponding functionals have the form (2.1). Thus, we recall the definitions of various functionals, explain that they represent special cases of (2.1), and overview possible available results on uniqueness or symmetry. We present also available results on the ability of the functionals to capture the symmetry of the underlying distribution. If not stated otherwise, we were not able to find any such result. To the best of our knowledge, there are no available corresponding results for tests of symmetry.

In this paper, functionals corresponding to various estimators will be considered. Functionals are defined for a distribution F in \mathbb{R}^p , while the empirical distribution is denoted as F_n . The functional is a population counterpart of the corresponding estimate and replacing F by F_n in the definitions of functionals yields estimates which can be computed from data $\mathbf{X}_1, \dots, \mathbf{X}_n$.

The class of multivariate M-estimators with auxiliary scale (not be confused with other multivariate M-estimators) was proposed by Tatsuoka and Tyler (2000) together with the corresponding functionals. Let us now assume that a scale functional $\sigma(F) > 0$ is given. For the multivariate location and scatter, the functionals are defined as the pairs $(\mathbf{m}_M(F), \mathbf{\Sigma}_M(F))$ solving

$$\min_{(\mathbf{m}, \mathbb{G}) \in \Theta} E \rho \left(\frac{(\mathbf{X} - \mathbf{m})^T \mathbb{G}^{-1} (\mathbf{X} - \mathbf{m})}{\sigma^2(F)} \right) \quad \text{s.t. } \det(\mathbb{G}) = 1,$$

where ρ is a given function. Tatsuoka and Tyler (2000) assessed the uniqueness of M-functionals with auxiliary scale under some assumptions. They also claimed the functionals to be centrally symmetric, but under assumptions which are stronger compared to those of our [Theorem 1](#).

S-estimators of multivariate location and scatter were defined by Lopuhaä (1989) and the corresponding S-functionals were formulated by Tatsuoka and Tyler (2000). A given function ρ is assumed, which is defined for $s \geq 0$, is nondecreasing, continuous from above at zero, and fulfills $0 = \rho(0) < \rho(\infty) < \infty$. S-functionals are defined as the pairs $(\boldsymbol{\mu}_S(F), \mathbb{S}_S(F))$ solving

$$\min_{(\mathbf{m}, \mathbb{S}) \in \Psi} \det(\mathbb{S}) \quad \text{s.t. } E \rho \left(\sqrt{(\mathbf{X} - \mathbf{m})^T \mathbb{S}^{-1} (\mathbf{X} - \mathbf{m})} \right) = b_0$$

for a positive b_0 . S-estimators are special cases of M-estimators with auxiliary scale as shown in [Theorem 2.1](#) of Tatsuoka and Tyler (2000). Multivariate S-functionals are

uniquely defined only at unimodal elliptically symmetric distributions. This uniqueness was established by Lopuhaä (1989) and also by Davies (1987), where the latter used a slightly different version of S-functionals. Later, uniqueness of S-functionals was derived under broader (non-elliptical) classes of symmetric distributions by Tatsuoka and Tyler (2000). Other work was devoted to finding rules for choosing proper parameters of multivariate S-estimators (Rocke 1996). It was however shown by Lopuhaä (1989) that S-estimators cannot achieve small asymptotic variance and 50 % breakdown point simultaneously.

The minimum volume ellipsoid (MVE) estimator was proposed by Rousseeuw (1984) and corresponds to a multivariate S-estimator with a zero-one step function ρ . The MVE functional is a special case of multivariate S-functionals, as shown by Tatsuoka and Tyler (2000). The uniqueness of the MVE functional is obvious for unimodal elliptic distribution as claimed by He and Wang (1996). More general results on the uniqueness of multivariate S-functionals by Lopuhaä (1989) and later by Tatsuoka and Tyler (2000) are valid also for the MVE functional.

Constrained M-estimators (CM-estimators) were proposed by Kent and Tyler (1996). CM-functionals for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined as pairs $(\boldsymbol{\mu}_{CM}(F), \boldsymbol{\Sigma}_{CM}(F))$ solving

$$\begin{aligned} \min_{(\boldsymbol{m}, \mathbb{S}) \in \Theta} & \mathbb{E} \left[\rho \left((\boldsymbol{X} - \boldsymbol{m})^T \mathbb{S}^{-1} (\boldsymbol{X} - \boldsymbol{m}) \right) \right] + \frac{1}{2} \log \det(\mathbb{S}) \\ \text{s.t.} & \mathbb{E} \left[\rho \left((\boldsymbol{X} - \boldsymbol{m})^T \mathbb{S}^{-1} (\boldsymbol{X} - \boldsymbol{m}) \right) \right] \leq \varepsilon \rho(\infty), \end{aligned}$$

where $\rho(s)$ is a bounded nondecreasing function for $s \geq 0$ and ε is a fixed value between 0 and 1. CM-estimators are special cases of M-estimators with auxiliary scale as shown in Theorem 2.1 of Tatsuoka and Tyler (2000). Uniqueness of CM-functionals was derived by Kent and Tyler (2001) under some assumptions. Their uniqueness under non-elliptical distributions was investigated by Tatsuoka and Tyler (2000). Kent and Tyler (1996) presented the following symmetry property as a by-product. If the distribution $\mathcal{L}(\boldsymbol{X})$ is centrally symmetric, then the CM-functional coincides with the center of symmetry.

Multivariate MM-estimators were proposed by Tatsuoka and Tyler (2000). Salibián-Barrera, Van Aelst, and Willems (2006) proposed multivariate MM-functionals using $\sigma_S(F)$ as scale of an S-functional with some function ρ_0 and using a function ρ_1 fulfilling their assumptions (R1) and (R2); the MM-functionals of location and shape $(\boldsymbol{\mu}_{MM}(F), \boldsymbol{\Gamma}_{MM}(F))$ are then defined as argument of

$$\min_{(\boldsymbol{m}, \mathbb{G}) \in \Theta} \mathbb{E} \rho_1 \left(\frac{\sqrt{(\boldsymbol{X} - \boldsymbol{m})^T \mathbb{G}^{-1} (\boldsymbol{X} - \boldsymbol{m})}}{\sigma_S(F)} \right) \quad \text{s.t.} \quad \det(\mathbb{G}) = 1$$

and the MM-functional for the covariance matrix as $\boldsymbol{\Sigma}_{MM}(F) = \sigma_S^2(F) \boldsymbol{\Gamma}_{MM}(F)$. Lopuhaä (1992) proposed an alternative version of MM-functionals only for $\boldsymbol{\mu}$, while the covariance matrix must be estimated beforehand. Because MM-estimators lie within the class of M-estimators with auxiliary scale, the results on uniqueness of Tatsuoka and Tyler (2000) are valid also for them. Lopuhaä (1992) claimed that if the distribution $\mathcal{L}(\boldsymbol{X})$ is elliptically contoured, the MM-functional for location coincides with the center of symmetry.

The minimum weighted covariance determinant (MWCD) estimator was proposed by Roelant, Van Aelst, and Willems (2009). The MWCD-functionals of location and shape are defined as any pair $(\boldsymbol{\mu}_{MWCD}(F), \mathbb{I}_{MWCD}(F))$ which gives the argument of

$$\min_{(\mathbf{m}, \mathbb{G} \in \Theta)} \mathbb{E} \left[h^+ \left(G \left(d_x^2(\mathbf{m}, \mathbb{G}) \right) \right) d_x^2(\mathbf{m}, \mathbb{G}) \right] \quad \text{s.t. } \det(\mathbb{G}) = 1,$$

where $G(t) = P(d_x^2(\mathbf{m}, \mathbb{G}) < t)$, $d_x^2(\mathbf{m}, \mathbb{G}) = (\mathbf{x} - \mathbf{m})^T \mathbb{G}^{-1} (\mathbf{x} - \mathbf{m})$, and $h^+ : (0, 1) \rightarrow [0, \infty)$ is a weight function which fulfills $\sup\{u; h^+(u) > 0\} = 1 - \alpha$ with $0 \leq \alpha \leq 1/2$ and $h^+(u) > 0$ for $u \in (0, 1 - \alpha]$. The MWCD functional for δ is obtained subsequently as $\boldsymbol{\Sigma}_{MWCD}(F) =$

$$= c_{h^+} \frac{\int \left[h^+ \left(G \left(d_x^2(\boldsymbol{\mu}_{MWCD}(F), \mathbb{G}_{MWCD}(F)) \right) \right) (\mathbf{x} - \boldsymbol{\mu}_{MWCD}(F)) (\mathbf{x} - \boldsymbol{\mu}_{MWCD}(F))^T \right] dF(\mathbf{x})}{\int h^+ \left(G \left(d_x^2(\boldsymbol{\mu}_{MWCD}(F), \mathbb{G}_{MWCD}(F)) \right) \right) dF(\mathbf{x})}$$

and MWCD functionals are uniquely defined under elliptically symmetric unimodal distributions, which follows from their Fisher consistency at such distributions as proven by Roelant, Van Aelst, and Willems (2009).

The minimum covariance determinant (MCD) estimator, proposed by Rousseeuw and van Driessen (1999), can be described as a special case of the MWCD with the weight function

$$h^+(u) = \mathbb{1}[u > k] \quad \text{for some } \frac{n}{s} \leq k \leq n,$$

where $\mathbb{1}$ is an indicator function. The corresponding MCD-functional was investigated by Cator and Lopuhaä (2012).

The MCD estimator fulfills (2.1), which follows from its being a special case of the MWCD estimator as explained by Agulló, Croux, and Van Aelst (2008) or Roelant, Van Aelst, and Willems (2009). Properties of the MCD estimator were overviewed by Hubert and Debruyne (2010) and uniqueness of the MCD functional was proven by Butler, Davies, and Jhun (1993) for distributions that have a unimodal elliptically contoured density.

The class of multivariate τ -estimators, which are able to combine a high robustness (breakdown point and bounded influence) with good asymptotic efficiency, and their corresponding τ -functionals was proposed by Lopuhaä (1991). τ -functionals of location and shape are defined as the pairs $(\boldsymbol{\mu}_\tau(F), \mathbb{I}_\tau(F))$ solving

$$\begin{aligned} \min_{(\mathbf{m}, \mathbb{G} \in \Theta)} \det(\mathbb{G}) \cdot \left[\mathbb{E} \rho_2 \left(\sqrt{(\mathbf{X} - \mathbf{m})^T \mathbb{G}^{-1} (\mathbf{X} - \mathbf{m})} \right) \right]^p \\ \text{s.t. } \mathbb{E} \rho_1 \left((\mathbf{X} - \mathbf{m})^T \mathbb{G}^{-1} (\mathbf{X} - \mathbf{m}) \right) = b_1 \end{aligned}$$

for given nonnegative functions ρ_1 and ρ_2 and a positive b_1 . Subsequently, the τ -functional for the covariance matrix is defined as

$$\boldsymbol{\Sigma}_\tau(F) = \frac{1}{b_2} \mathbb{I}_\tau(F) \mathbb{E} \rho_2 \left(\sqrt{(\mathbf{X} - \boldsymbol{\mu}_\tau(F))^T \mathbb{I}_\tau^{-1}(F) (\mathbf{X} - \boldsymbol{\mu}_\tau(F))} \right),$$

where $b_2 > 0$ is a given constant. τ -functionals are uniquely defined for elliptical distributions if using a suitable value of b_1 , as shown by Lopuhaä (1991).

An L_1 -type estimator of multivariate location and shape was proposed by Roelant and van Aelst (2007). Location and shape are defined simultaneously and the corresponding functionals are defined as the pair $(\boldsymbol{\mu}_1(F), \boldsymbol{\Pi}_1(F))$ which solves

$$\min_{(\mathbf{m}, \mathbb{G}) \in \Theta} E \sqrt{(\mathbf{X} - \mathbf{m})^T \mathbb{G}^{-1} (\mathbf{X} - \mathbf{m})} \quad \text{s.t. } \det(\mathbb{G}) = 1.$$

The estimator extends the concept of the univariate median and belongs to the class of multivariate M-estimators. The uniqueness of the functional for location and shape is ensured. Roelant and van Aelst (2007) also formulated the covariance matrix functional based on $(\boldsymbol{\mu}_1(F), \boldsymbol{\Pi}_1(F))$, which however does not seem to fulfill (2.1).

Elliptical quantiles for multivariate data proposed by Hlubinka and Šiman (2013) also preserve the centers and axes of symmetry, just like the nonlinear versions of the quantiles by Hlubinka and Šiman (2015).

Let us consider also the standard estimates in the form of the mean and empirical covariance matrix, if estimated simultaneously. The corresponding functionals, i.e. the expectation and population covariance matrix, fulfill (2.1), because the pair is a special case of some of the functionals described above. They are a unique solution of the corresponding optimization task (2.1) and [Theorem 1](#) holds trivially for them.

In addition, other interesting multivariate estimation approaches include the symmetrized M-estimators of multivariate scatter of Sirkiä, Taskinen, and Oja (2007), which are (as functionals) diagonal if the components of the random vector are independent, or the multivariate Forward Search (exploiting weighted versions of standard estimators), whose strong consistency at multivariate normal models and high breakdown point was derived by Cerioli, Farcomeni, and Riani (2014).

4. Simulations

The aim of the simulations is to investigate the performance of the newly proposed tests under different conditions and to compare them for different estimators. The null hypothesis of interest is symmetry of the underlying distribution around all three coordinate axes. For various situations with $p = 3$, we compute various estimators of multivariate parameters described in this paper and consider tests of the null hypothesis that $\boldsymbol{\Sigma}$ (which exists in all examples) is diagonal against a general alternative hypothesis that H_0 does not hold.

We compare the performance of tests based on various estimators $\hat{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$. Denoting elements of $\hat{\boldsymbol{\Delta}}$ by $\hat{\Delta}_{ij}$, nonparametric combination of three tests based on $\hat{\Delta}_{12}, \hat{\Delta}_{13}$ and $\hat{\Delta}_{23}$ jointly are exploited. The test is described in [Algorithm 1](#), inspired by the general algorithm of [Bonnini et al. \(2014\)](#), using one of these combination functions as a special case:

- Fisher $\psi(\lambda_1, \lambda_2, \lambda_3) = -2 \sum_k \log(\lambda_k)$;
- Liptak $\psi(\lambda_1, \lambda_2, \lambda_3) = \sum_k \Phi^{-1}(1 - \lambda_k)$;
- Tippett $\psi(\lambda_1, \lambda_2, \lambda_3) = \max_k \{1 - \lambda_k\}$.

All computations were performed in *R* software (R Core Team 2018), exploiting the *rrcov* package of Todorov and Filzmoser (2009) for robust estimators. Besides from the classical estimators (i.e. the mean and empirical covariance matrix), we consider the MCD and MVE (both taking the optimization criterion over $\lfloor n/2 \rfloor$ observations), S-estimators with breakdown point 0.5, and MM-estimators with breakdown point 0.5 and efficiency 0.95.

Simulations were performed for data generated from five following models with $n = 80$, where the data allow to reveal the effect of moving away from H_0 . In each case, we repeat Algorithm 1 for 10 000 randomly generated datasets with the choice $B = 1000$.

Algorithm 1. Test of Section 4 based on the nonparametric combination methodology

Input: Data $\mathbf{X} \in \mathbb{R}^{n \times p}$ with $p = 3$, selected combination function ψ , constant $B > 0$

Input: Multivariate estimator fulfilling Theorem 1, whose elements of Δ obtained from \mathbf{X} are denoted as $\hat{\Delta}_{ij}(\mathbf{X})$, where $i, j = 1, 2, 3$

Output: The combined p -value λ'

- 1: Compute the test statistic denoted as $T = (T_1, T_2, T_3)^T := (\hat{\Delta}_{12}(\mathbf{X}), \hat{\Delta}_{13}(\mathbf{X}), \hat{\Delta}_{23}(\mathbf{X}))^T$
- 2: **for** $b = 1$ to B **do**
- 3: Generate independent random variables $G_{11}, \dots, G_{np} \in \mathbb{R}^{n \times p}$, where $P(G_{kl} = 1) = 1/2$ and $P(G_{kl} = -1) = 1/2$ for each $k = 1, \dots, n$ and $l = 1, \dots, p$
- 4: Consider the data (say) $\mathbf{X}_{(b)}^*$ with elements $X_{kl(b)}^* = X_{kl}G_{kl}$ for each $k = 1, \dots, n$ and $l = 1, \dots, p$
- 5: Compute the corresponding test statistic

$$T_{(b)}^* = (T_{1(b)}^*, T_{2(b)}^*, T_{3(b)}^*)^T := \left(\hat{\Delta}_{12}(\mathbf{X}_{(b)}^*), \hat{\Delta}_{13}(\mathbf{X}_{(b)}^*), \hat{\Delta}_{23}(\mathbf{X}_{(b)}^*) \right)^T$$

6: **end for**

7: Denote $\hat{L}_s(z) = \{1/2 + \sum_{b=1}^B \mathbb{1}[T_{s(b)}^* \geq z]\} / (B + 1)$ for $s = 1, 2, 3$

8: Compute $\lambda_{s(b)}^* = \hat{L}_s(T_{s(b)}^*)$ and $\lambda_s = \hat{L}_s(T_s)$ for each $b = 1, \dots, B$ and $s = 1, 2, 3$

9: Compute $T' = \psi(\lambda_1^*, \lambda_2^*, \lambda_3^*)$ and $T'_{(b)} = \psi(\lambda_{1(b)}^*, \lambda_{2(b)}^*, \lambda_{3(b)}^*)$ for each b

10: Compute $\lambda' = \sum_b \mathbb{1}[T'_{(b)} \geq T'] / B$

- A. 3-dimensional normal distribution $N_3(0, \mathbf{\Delta}_0)$ with $\mathbf{\Delta}_0 = (1-d)\mathcal{I} + d\mathbf{e}\mathbf{e}^T$, where \mathcal{I} is a unit matrix of size 3×3 and $\mathbf{e} = (1, 1, 1)^T$.
- B. Data are created as in study A and contaminated by 5 gross outliers; there are 5 randomly selected values, which are replaced by randomly generated values from $N_3(0, 5\mathbf{\Delta}_0)$. Again, values of d ranging from 0 to 0.9 are used. Because of the symmetric contamination, $d = 0$ corresponds to H_0 .
- C. 3-dimensional vector $(X_1, X_2, X_3)^T$ with independent components, where X_1 is generated from $N(0, 0.36)$, X_2 from uniform on $[-2, 2]$, and X_3 from Laplace distribution with $EX_3 = 0$ and $\text{var } X_3 = 2$. Then, the data are rotated about the

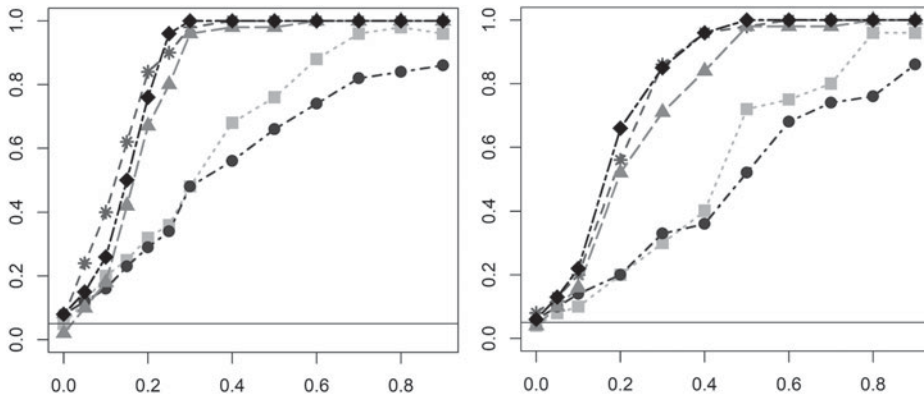


Figure 1. Results of simulation study A (left) and B (right). Tests based on classical estimates (stars), MVE (dark circles), MCD (light squares), S-estimator (light triangles), and MM-estimator (dark diamonds).

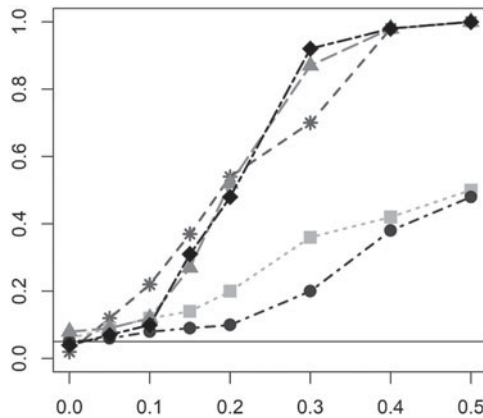


Figure 2. Results of simulation study C. Tests based on classical estimates (stars), MVE (dark circles), MCD (light squares), S-estimator (light triangles), and MM-estimator (dark diamonds).

third axis by an angle θ ranging from 0 to 0.5 (in radians), where $\theta=0$ corresponds to H_0 .

- D. 3-dimensional vector $(X_1, X_2, X_3)^T$ with independent components, where X_1 , X_2 and X_3 are generated from t_2 distribution, but X_3 is replaced by $X_3/5$. Then, the data are rotated about the third axis by an angle θ ranging from to 0.5 (in radians), where $\theta=0$ corresponds to H_0 .
- E. Data are created as in study D and contaminated by 4 gross outliers; these are 4 randomly selected values, which are replaced by values $(10 + Z_1, 5 + Z_2, 0)^T$, where $Z_1 \sim N(0, 0.25)$ and $Z_2 \sim N(0, 0.25)$ are independent random variables.

The results are presented as power curves in [Figures 1](#) (studies A and B), [2](#) (study C), and [3](#) (studies D and E). For studies A to D, the tests hold the probability of type I error at the 5 % level, due to their construction (property of nonparametric combination methodology), so the horizontal line corresponding to 0.05 is shown in the figures. The power of the tests increases as the data become more distant from H_0 . In

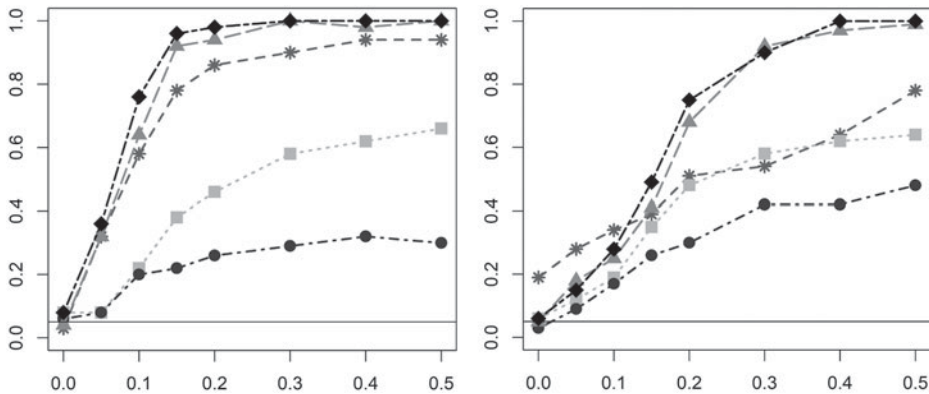


Figure 3. Results of simulation study D (left) and E (right). Tests based on classical estimates (stars), MVE (dark circles), MCD (light squares), S-estimator (light triangles), and MM-estimator (dark diamonds).

study E, no choice of θ corresponds to H_0 due to the asymmetric contamination; however, if the outliers were ignored, $\theta = 0$ would correspond to H_0 .

Comparing the individual estimates, tests based on the classical ones yield the largest power in study A and partially in study D (only for small deviations from H_0). Tests based on MM-estimation yield the largest power in the remaining studies. Study E advocates using the tests based on robust estimates, which have the level close to 0.05, while the test based on classical estimates has the level equal to 0.19. Tippett's method outperforms the other nonparametric combination approaches in studies A and B, and Fisher's does the same in studies C, D and E. We do not present results for a rotation by larger angles in the studies C, D and E, as the data turn back to H_0 and the powers decrease there. Additional results (not presented here) obtained with tests based on MM-estimators with a lower efficiency (below 0.95) yield weaker results compared to those of MM-estimators presented here.

5. A real data example

We illustrate the performance of the proposed approach on the Australian athletes dataset, which is a real dataset publicly available e.g. in the R software package DAAG (Maindonald and Braun 2015). We consider three variables, particularly the red blood cell count (denoted here as X_1), white blood cell count (X_2), and hemoglobin concentration (X_3). The dataset with $n = 202$ measurements on (both male and female) athletes was previously investigated by Henze, Hlávka, and Meintanis (2014), who tested spherical symmetry (and clearly rejected it).

We are interested in testing the symmetry of the data around all three coordinates axes (after centering the data). As the correlation coefficient between X_1 and X_3 is large ($r = 0.889$), the overall null hypothesis of this symmetry seems very unlikely; the other correlations equal to $r(X_1, X_2) = 0.147$ and $r(X_2, X_3) = 0.135$. Algorithm 1 is used for the testing again with $B = 1000$. Table 1 gives p -values of individual (i.e. permutation) tests for individual symmetries around an axis, and the p -value of the overall test based on the nonparametric combination. Every test rejects H_0 , while tests based on the MCD

Table 1. Results of permutation tests in the Australian athletes dataset for the whole dataset (left) and averaged results for 1000 random subsamples with $n = 50$. The p -values for the three individual tests are given together with the p -value obtained by the Fisher's (F), Liptak's (L) and Tippett's (T) combination method.

Estimator	The whole dataset						Random subsamples with $n = 50$					
	Individual tests			Overall test			Individual tests			Overall test		
	1-2	1-3	2-3	F	L	T	1-2	1-3	2-3	F	L	T
Clas.	0.03	0.00	0.07	0.00	0.00	0.00	0.68	0.01	0.36	0.01	0.04	0.01
MCD	0.02	0.00	0.02	0.00	0.00	0.00	0.58	0.01	0.34	0.04	0.09	0.07
MVE	0.38	0.21	0.18	0.35	0.24	0.29	0.83	0.35	0.38	0.67	0.26	0.39
S	0.03	0.00	0.05	0.00	0.00	0.00	0.69	0.01	0.32	0.01	0.05	0.01
MM	0.02	0.00	0.04	0.00	0.00	0.00	0.77	0.01	0.34	0.01	0.05	0.01

seem to be the most powerful. In addition, we consider 1000 random subsamples selected from the dataset with $n = 50$ measurements. The averaged powers of the tests shown in right part of Table 1 are reduced compared to those over the whole dataset. Still, we may conclude also for the subsamples that the symmetry around all three axes is clearly rejected.

6. Conclusions

This paper investigates symmetry aspects of functionals corresponding to various multivariate location and scatter estimators, particularly their ability to capture the symmetry of the underlying distribution. Various forms of symmetry considered in the paper include central symmetry, marginal symmetry, symmetry around an affine subspace, and symmetry around a coordinate axis. Such various forms of symmetry are useful e.g. as assumptions of various economic models.

The contribution of the paper is a formulation of very general sufficient conditions, which ensure various symmetry properties. Examples of robust multivariate estimators, which fulfill these conditions, are discussed in detail in Sec. 3. The symmetry results are also valid for popular classes of multivariate estimators, including S-estimators, the MCD or MVE, MM-estimators or τ -estimators. A similar seems implicit in the work of Dümbgen, Pauly, and Schweizer (2015) for M-functionals of multivariate scatter.

To the best of our knowledge, the new tests are the first general symmetry tests for dimension $p > 2$, which are at the same time valid also for non-elliptical distributions. Simulations were performed to investigate the performance of the newly proposed tests. The tests of Sec. 2.3 based on Theorem 1 are expressed for the matrix \mathbb{A} and are valid also for the covariance matrix $\mathbb{\Sigma}$; we also consider estimators of $\mathbb{\Sigma}$ in the numerical studies of Sec. 4, where the performance of the new tests is illustrated. Tests based on the mean and empirical covariance matrix turn out to be the best choice for normally distributed multivariate data, but they are not able to keep the level under asymmetric contamination. Tests based on MM-estimators turn out to be the best solution for non-normal and/or contaminated data.

As a future work, it would be straightforward to formulate Theorem 1 also for the context of linear regression with a multivariate response. Then, analogous results are valid e.g. for multivariate S-estimators of van Aelst and Willems (2005) or the estimator of Ben, Martínez, and Yohai (2006).

Funding

The work is supported by the grant “Nonparametric (statistical) methods in modern econometrics” No. 17-07384S of the Czech Science Foundation. The author would like to thank Miroslav Šiman and the anonymous referees for valuable suggestions.

ORCID

Jan Kalina  <http://orcid.org/0000-0002-8491-0364>

References

- Van Aelst, S., and G. Willems. 2005. Multivariate regression S-estimators for robust estimation and inference. *Statistica Sinica* 15:981–1001.
- Agulló, J., C. Croux, and S. Van Aelst. 2008. The multivariate least-trimmed squares estimator. *Journal of Multivariate Analysis* 99 (3):311–38. doi:10.1016/j.jmva.2006.06.005.
- Baringhaus, L. 1991. Testing for spherical symmetry of a multivariate distribution. *The Annals of Statistics* 19 (2):899–917. doi:10.1214/aos/1176348127.
- Ben, M. G., E. Martínez, and V. J. Yohai. 2006. Robust estimation for the multivariate linear model based on a τ -scale. *Journal of Multivariate Analysis* 97 (7):1600–22. doi:10.1016/j.jmva.2005.08.007.
- Bonnini, S., L. Corain, M. Marozzi, and L. Salmaso. 2014. *Nonparametric hypothesis testing: Rank and permutation methods with applications in R*. New York: Wiley.
- Butler, R. W., P. L. Davies, and M. Jhun. 1993. Asymptotic for the minimum covariance determinant estimator. *The Annals of Statistics* 21 (3):1385–400. doi:10.1214/aos/1176349264.
- Cator, E. A., and H. P. Lopuhaä. 2012. Central limit theorem and influence function for the MCD estimators at general multivariate distributions. *Bernoulli* 18 (2):520–51. doi:10.3150/11-BEJ353.
- Ceroli, A., A. Farcomeni, and M. Riani. 2014. Strong consistency and robustness of the forward search estimator of multivariate location and scatter. *Journal of Multivariate Analysis* 126: 167–83. doi:10.1016/j.jmva.2013.12.010.
- Davies, P. L. 1987. Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics* 15 (3):1269–92. doi:10.1214/aos/1176350505.
- Dümbgen, L., M. Pauly, and T. Schweizer. 2015. M-functionals of multivariate scatter. *Statistics Surveys* 9:32–105. doi:10.1214/15-SS109.
- Dutta, S., A. K. Ghosh, and P. Chaudhuri. 2011. Some intriguing properties of Tukey’s half-space depth. *Bernoulli* 17 (4):1420–34. doi:10.3150/10-BEJ322.
- García-Portugués, E., D. Paindaveine, and T. Verdebout. 2018. On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *ArXiv* 1706:05030.
- He, X., and G. Wang. 1996. Cross-checking using the minimum volume ellipsoid estimator. *Statistica Sinica* 6:367–74.
- Henze, N., Z. Hlávka, and S. G. Meintanis. 2014. Testing for spherical symmetry via the empirical characteristic function. *Statistics* 48:1282–96.
- Hlubinka, D., and M. Šiman. 2013. On elliptical quantiles in the quantile regression setup. *Journal of Multivariate Analysis* 116:161–71.
- Hlubinka, D., and M. Šiman. 2015. On generalized elliptical quantiles in the nonlinear quantile regression setup. *Test* 24 (2):249–64. doi:10.1007/s11749-014-0405-3.
- Hubert, M., and M. Debruyne. 2010. Minimal covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (1):36–43. doi:10.1002/wics.61.
- Kent, J. T., and D. E. Tyler. 1996. Constrained M-estimation for multivariate location and scatter. *The Annals of Statistics* 24 (3):1346–70. doi:10.1214/aos/1032526973.
- Kent, J. T., and D. E. Tyler. 2001. Regularity and uniqueness for constrained M-estimates and redescending M-estimates. *The Annals of Statistics* 29(1):252–65. doi:10.1214/aos/996986508.

- Lopuhaä, H. P. 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance. *Annals of Statistics* 17:1662–83.
- Lopuhaä, H. P. 1991. Multivariate τ -estimators for location and scatter. *Canadian Journal of Statistics* 19 (3):307–21. doi:10.2307/3315391.n.
- Lopuhaä, H. P. 1992. Highly efficient estimators of multivariate location with high breakdown point. *Annals of Statistics* 20:398–413.
- Maindonald, J. H., and W. J. Braun. 2015. DAAG: Data analysis and graphics data and functions. R package version 1.22, <https://CRAN.R-project.org/package=DAAG>.
- Oja, H. 2010. Multivariate nonparametric methods with R: An approach based on spatial signs and ranks. In *Lecture notes in statistics*, vol. 199. New York: Springer.
- Paindaveine, D., and G. Van Bever. 2013. From depth to local depth: A focus on centrality. *Journal of the American Statistical Association* 108 (503):1105–19. doi:10.1080/01621459.2013.813390.
- Pesarin, F., and L. Salmaso. 2010. *Permutation tests for complex data: Theory, applications and software*. New York: Wiley.
- Quessy, J. F. 2016. On consistent nonparametric statistical tests of symmetry hypotheses. *Symmetry* 8 (5):31. doi:10.3390/sym8050031.
- R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rao, K. S., and M. Raghunath. 2012. A simple nonparametric test for bivariate symmetry about a line. *Journal of Statistical Planning and Inference* 142:430–44. doi:10.1016/j.jspi.2011.07.025.
- Rocke, D. M. 1996. Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of Statistics* 24 (3):1327–45. doi:10.1214/aos/1032526972.
- Roelant, E., and S. Van Aelst. 2007. An L_1 -type estimator of multivariate location and shape. *Statistical Methods and Applications* 15 (3):381–93. doi:10.1007/s10260-006-0030-8.
- Roelant, E., S. Van Aelst, and G. Willems. 2009. The minimum weighted covariance determinant estimator. *Metrika* 70 (2):177–201. doi:10.1007/s00184-008-0186-3.
- Roelant, E., S. Van Aelst, and C. Croux. 2009. Multivariate generalized S-estimators. *Journal of Multivariate Analysis* 100 (5):876–87. doi:10.1016/j.jmva.2008.09.002.
- Rousseeuw, P. J. 1984. Least median of squares regression. *Journal of the American Statistical Association* 79 (388):871–80. doi:10.1080/01621459.1984.10477105.
- Rousseeuw, P. J., and K. Van Driessen. 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 34:212–23. doi:10.1080/00401706.1999.10485670.
- Rousseeuw, P. J., and A. Struyf. 2002. A depth test for symmetry. In *Goodness-of-fit tests and model validity*, eds. C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, and M. Mesbah, 401–412. New York: Springer Science + Business Media.
- Salibián-Barrera, M., S. Van Aelst, and G. Willems. 2006. Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association* 101 (475):1198–211. doi:10.1198/016214506000000096.
- Serfling, R. 2006. Multivariate symmetry and asymmetry. In *Encyclopedia of statistical sciences*, eds. S. Kotz, N. Balakrishnan, C. B. Read, and B. Vidakovic, 2nd ed., vol. 8, 5338–5345. New York: Wiley.
- Sirkiä, S., S. Taskinen, and H. Oja. 2007. Symmetrised M-estimators of multivariate scatter. *Journal of Multivariate Analysis* 98 (8):1611–29. doi:10.1016/j.jmva.2007.06.005.
- Tatsuoka, K. S., and D. E. Tyler. 2000. On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *The Annals of Statistics* 28:1219–43. doi:10.1214/aos/1015956714.
- Todorov, V., and P. Filzmoser. 2009. An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software* 32 (3):1–47.