Discussiones Mathematicae
Probability and Statistics 36 (2016) 53–66
doi:10.7151/dmps.1184

ROBUST ESTIMATION IN THE MULTIVARIATE NORMAL MODEL

Agnieszka Kulawik

Institute of Mathematics
University of Silesia
40–007 Katowice, Bankowa 14, Poland

e-mail: Agnieszka.Kulawik@us.edu.pl

and

Stefan Zontek

Faculty of Mathematics, Computer Science and Econometrics
University of Zielona Góra
65–516 Zielona Góra, Szafrana 4A, Poland

e-mail: S.Zontek@wmie.uz.zgora.pl

Abstract

Robust estimation presented in the following paper is based on Fisher consistent and Fréchet differentiable statistical functionals. The method has been used in the multivariate normal model with variance components [5]. To transfer the method to estimate vector of expectations and positive definite covariance matrix of the multivariate normal model it is required to express the covariance matrix as a linear combination of basic elements of the vector space of real, square and symmetric matrices. The theoretical results have been completed with computer simulation studies. The robust estimator has been investigated both for model and contaminated data. Comparison with the maximum likelihood estimator has also been included.

Keywords: asymptotic normality, Fisher consistency, Fréchet differentiability, multivariate normal model, statistical functional.

2010 Mathematics Subject Classification: 62H10, 62H12.

1. Introduction

The problem of robust estimation in multivariate case has been already investigated by many authors. One of the most popular work in this topic is Maronna's article [6]. Maronna's proposition based on solutions of properly defined system of equations. Another author that considered the problem is Rousseeuw [7]. He has defined minimum covariance determinant estimator (MCD estimator) using the subset of observations for which covariance matrix has the lowest determinant. Next proposition comes from Bednarski and Zontek [2]. They proposed a method of robust estimation of parameters in two way crossed classification model. The method based on Fréchet differentiable statistical functionals and involved Huber's function [3]. In the two way crossed classification model Bednarski and Zontek reduced the estimation problem to a number of simultaneous one dimensional estimation problems. To do this they represented the quadratic form in loglikelihood function as a sum of quadratic terms and used a smoothed version of Huber's proposition for scale estimation. Both in paper [8] of Zmyślony and Zontek, and in paper [5] of Kulawik and Zontek, the Bednarski and Zontek's method has been adapted. Zmyślony and Zontek considered the variance components model with commutative covariance matrices. Kulawik and Zontek applied the method in any variance components model. In this paper we have shown that such kind of modification can be applied also in multivariate normal model. In case of the variance components models covariance matrix is a finite generated convex cone of form $\sum_{i=1}^{k} \sigma_i^2 V_i$, where matrices $V_1, \ldots, V_k$ are known, linearly independent, symmetric, nonnegative-definite and $\sum_{i=1}^{k} V_i$ is positive-definite. Scale parameters $\sigma_1, \ldots, \sigma_k$ were estimated using basic computers procedures. In case of the multivariate normal models covariance matrix has form $\sum_{i=1}^{k} \alpha_i W_i$, where $W_1, \ldots, W_k$ are elements of a basis of the space of real, square and symmetric matrices. Using basic computer procedures for estimating $\alpha_1, \ldots, \alpha_k$, the matrix $\hat{\Sigma} = \sum_{i=1}^{k} \hat{\alpha}_i W_i$ could be negative-definite, hence $\tilde{\Sigma} = (\hat{\Sigma} \cdot \hat{\Sigma})^{1/2}$ has been involved to avoid the problem. It means that basic computer procedures can be used also in the case of the multivariate normal model.

Our way to obtain robust estimators of parameters of the multivariate normal model gives similar conditions to that elaborated by Maronna [6]. We are concentrated on two properties of statistical functionals defining robust estimators: Fisher consistency and Fréchet differentiability under normal model. We give formula for the asymptotic covariance matrix not only for estimator of the expected value but for estimator of the covariance matrix, too.

The paper consists of four sections, references and appendix. The first section is a brief introduction to the subject. In the second one some basic information on statistical functionals are presented. There can be also found a connection between the statistical functional used in the paper and Clarke's [1] one. The third section gives a form of our robust estimator for unknown both shift parameter and positive-definite covariance matrix of the multivariate normal model. There have been also shown asymptotic properties of the estimator. In the fourth section results of computer simulation are included. In the computer studies the maximum likelihood estimator (m.l.e.) and our robust estimator are compared both for model and contaminated data. It occurs that the robust estimator is sensible for model data and also obvious superiority over m.l.e. for contaminated data is shown. After references comes appendix. The appendix consists of some technical lemmas used in the paper.

2. Basic information and notation

Let us introduce some notation. Assume that for $r, s \in \mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{R}^r_s$ denotes the set of all real matrices with r rows and s columns, I_r denotes the identity matrix of size r and $0^r = (0, \dots, 0)^T \in \mathbb{R}^r$. Moreover let $\mathrm{tr}(A)$ and $|A|$ stand for trace and determinant of matrix A, respectively, and let $A > 0$ means that A is positive definite. Furthermore $\mathrm{N}_r(\mu, \Sigma)$ will denote multivariate (r-variate) normal distribution with mean vector μ and positive-definite covariance matrix Σ.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\Theta \subset \mathbb{R}^m$ be a parameter space for some $m \in \mathbb{N}$. Assume that $J \in \mathbb{N}$. We consider the statistical model $\{F_\theta : \theta \in \Theta\}$ — a subset of the family $\mathcal{G}$ (if $F \in \mathcal{G}$, then $F\colon \mathbb{R}^J \to \mathbb{R}$ is a distribution function).

Definition. A function defined on a nonempty subset of $\mathcal{G}$ and taking values in Θ is called *a statistical functional* (also *von Mises functional*).

If T is a statistical functional and $\hat{F}_N$, $N \in \mathbb{N}$, is the empirical distribution function of a random sample of size N drawn from F_θ, then $T(\hat{F}_N)$ is an estimator of parameter θ.

Example 1. M.l.e. of parameter $\theta = (\mu, \Sigma)$ of distribution $\mathrm{N}_J(\mu, \Sigma)$ can be defined by the statistical functional

$$\tilde{T}(G) = \mathrm{argmin}_{\theta \in \Theta} \int \left(\ln |\Sigma|^{\frac{1}{2}} + \frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu) \right) \mathrm{d}G(y), \tag{1}$$

for $G \in \mathcal{G}$, if the above integral is finite.

Example 2. Let $\Psi\colon \mathbb{R}^J \times \Theta \to \mathbb{R}^m$. A function

$$G \longrightarrow \mathrm{sol}\left\{ \int \Psi(y|\theta)\, \mathrm{d}G(y) = 0^m \right\}, \tag{2}$$

where "sol" means a properly chosen root of the equation in brackets, is a statistical functional, if there exists $G_0 \in \mathcal{G}$ such that equation $\int \Psi(y|\theta)\,\mathrm{d}G_0(y) = 0^m$ has a root in Θ.

Robust estimation considered in this paper is based on Fisher consistent and Fréchet differentiable statistical functionals.

Definition. Statistical functional T is *Fisher consistent* if $T(F_\theta) = \theta$ for all $\theta \in \Theta$.

Let us consider $\mathcal{G}$ with supremum metric $||\cdot||_\infty$, i.e., $||F-G||_\infty = \sup_{y\in\mathbb{R}^J}|F(y)-G(y)|$, where $|\cdot|$ is euclidean norm on $\mathbb{R}$. Denote by $\eta(\varepsilon, G)$ the neighborhood of size $\varepsilon > 0$ of $G \in \mathcal{G}$ defined by $\eta(\varepsilon, G) = \{F \in \mathcal{G} : ||F-G||_\infty < \varepsilon\}$. Let $\mathcal{D}$ denote the linear space spanned by the differences $F - G$ for all $F, G \in \mathcal{G}$.

Definition. Statistical functional T is *Fréchet differentiable at* $F \in \mathcal{G}$, if there exists a linear functional $T' \colon \mathcal{D} \to \mathbb{R}^m$ such that

$$|T(G) - T(F) - T'(G-F)| = o(||F-G||_\infty),$$

where $|\cdot|$ is euclidean norm on $\mathbb{R}^m$ and $\frac{o(||F-G||_\infty)}{||F-G||_\infty} \to 0$, while $||F-G||_\infty \to 0$.

In the Clarke's paper [1] there can be found a set of conditions implying Fréchet differentiability at F_θ, $\theta \in \Theta$, of statistical functional (2). Under the conditions linear functional T' has the form

$$T^{'}(G - F_\theta) = -M(\theta)^{-1} \int \Psi(y|\theta)\,\mathrm{d}(G - F_\theta)(y), \tag{3}$$

where

$$M(\theta) = \int \left[\frac{\partial \Psi(y|\theta)}{\partial \theta}\right] \mathrm{d}F_\theta(y).$$

Let $(G_N)_{N\in\mathbb{N}} \subset \mathcal{G}$. Assume that there exists $\delta > 0$ such that $G_N \in \eta(\frac{\delta}{\sqrt{N}}, F_\theta)$ for $N \in \mathbb{N}$.

Theorem 3. *Let $\hat{F_N}$ denote the empirical distribution function of a random sample $Y_1, \ldots, Y_N$ drawn from distribution G_N. If statistical functional T is Fisher consistent and Fréchet differentiable at F_θ and if linear functional T' has the form* (3), *then $\sqrt{N}(T(\hat{F}_N) - \theta)$ is asymptotically normal with mean vector* 0^m *and covariance matrix*

$$M(\theta)^{-1} \left\{ \int \Psi(y|\theta)\,[\Psi(y|\theta)]^T \,\mathrm{d}F_\theta(y) \right\} [M(\theta)^{-1}]^T.$$

Proof. A proof of the theorem is based on Kiefer's inequality [4] and central limit theorem (more details can be found in Kulawik and Zontek's article [5]). ∎

In the current paper statistical functional

$$T^*(G) = \operatorname{argmin}_{\theta\in\Theta} \int \Phi(y|\theta)\mathrm{d}G(y), \tag{4}$$

for $G \in \mathcal{G}$, will be considered,where $\Phi\colon \mathbb{R}^J \times \Theta \to \mathbb{R}$ is a function and if the term above exists. Function Φ is called *an objective function.* If $G_0 \in \mathcal{G}$, then $T^*(G_0)$ means that function

$$\theta \to \int \Phi(y|\theta)\mathrm{d}G_0(y)$$

attains global minimum for $\theta = T^*(G_0)$ and

$$\int \frac{\partial\Phi(y|\theta)}{\partial\theta}\mathrm{d}G_0(y) = 0^m$$

for $\theta = T^*(G_0)$. Hence $T^*(G_0)$ can be treated as the value of functional (2) for $G = G_0$ taking $\Psi(\cdot|\theta) = \frac{\partial\Phi(\cdot|\theta)}{\partial\theta}$ and assuming that $T^*(G_0)$ is a properly chosen root of the equation in brackets.

3. Robust estimation in the multivariate normal model

We consider a family of multivariate normal models $\mathrm{N}_J(X\beta, \Sigma)$ with an unknown vector $\beta \in \mathbb{R}^a$ and an unknown positive definite covariance matrix $\Sigma \in \mathbb{R}^J_J$. We assume that $X \in \mathbb{R}^J_a$ is a known matrix with a independent columns. Let matrices $W_1, \dots, W_k$, where $k = \frac{J(J+1)}{2}$, denote a basis of the vector space of real, square and symmetric matrices. Let $\alpha_1 \in \mathbb{R}, \dots, \alpha_k \in \mathbb{R}$ be coordinates of the matrix Σ with respect to the basis $W_1, \dots, W_k$, i.e., $\Sigma = \sum_{i=1}^k \alpha_i W_i$. The parameter $\theta = (\beta_1, \dots, \beta_a, \alpha_1, \dots, \alpha_k)^T = (\beta^T, \alpha^T)^T \in \Theta$ will be estimated, where

$$\Theta = \{(\beta_1, \dots, \beta_a, \alpha_1, \dots, \alpha_k)^T \in \mathbb{R}^{a+k} : \beta_1 \in \mathbb{R} \wedge \dots \wedge \beta_a \in \mathbb{R} \wedge \\ \alpha_1 \in \mathbb{R} \wedge \dots \wedge \alpha_k \in \mathbb{R} \wedge \sum_{i=1}^k \alpha_i W_i > 0\}.$$

We define statistical functional T^* of form (4) with objective function

$$\Phi(y|\theta) = \ln|\sum_{i=1}^k \alpha_i W_i|^{\frac{1}{2}} + \varphi\left[\frac{1}{c^2}(y - X\beta)^T \left(\sum_{i=1}^k \alpha_i W_i\right)^{-1} (y - X\beta)\right], \tag{5}$$

where $c > 0$ is a constant and $\varphi\colon [0, +\infty) \to \mathbb{R}$, to get an estimator of parameter θ.

Remark 4. For $\varphi(x) = \frac{1}{2}x$ and $c = 1$ we get the functional $\tilde{T}$ of form (1).

Fisher consistency and Fréchet differentiability at F_θ, $\theta \in \Theta$, of the functional T^* will be proved under the following conditions (B1)–(B4). The conditions are the same as in the case of multivariate normal model with variance components from paper of Kulawik and Zontek [5].

(B1) Function φ has positive derivative on $(0, +\infty)$.

(B2) Function $x\varphi'(x^2)$ has nonnegative derivative on $[0, +\infty)$ and there exists $x_0 > 0$ such that $2x_0^2\varphi'(x_0^2) > J$.

(B3) Function φ'' is continuous.

(B4) Functions $x\varphi'(x^2)$ and $x^2\varphi''(x^2)$ are bounded.

The constant c appearing in the definition of the objective function (5) has to be real positive number from the following lemma. It depends on both function φ and dimension J.

Lemma 5. *Let Z be a multivariate normally distributed random vector with mean vector 0^J and covariance matrix I_J. If $\varphi\colon [0, +\infty) \to \mathbb{R}$ satisfies* (B1) *and* (B2), *then there is a unique $c_{\varphi,J} > 0$ for which*

$$c \longmapsto J\ln(c) + \mathbb{E}\left[\varphi\left(\frac{Z^T Z}{c^2}\right)\right], \ c > 0, \tag{6}$$

attains the global minimum.

A proof of the lemma can be found in the paper of S. Zontek [9].
From now on we take $c = c_{\varphi,J}$ in definition of objective function (5). Now we can introduce the theorems concerning Fisher consistency and Fréchet differentiability of functional T^*. The following lemma will be used in proof of the first theorem.

Lemma 6. *Assume that $\Sigma_0 \in \mathbb{R}^J_J$ is positive-definite, symmetric and assume that $\mu_0 \in \mathbb{R}^J$. Let $F_{(\mu_0,\Sigma_0)}$ denote the distribution function of $\mathrm{N}_J(\mu_0, \Sigma_0)$. If $\varphi\colon [0, +\infty) \to \mathbb{R}$ satisfies* (B1) *and* (B2), *then*

$$\int \left[\ln|\Sigma|^{\frac{1}{2}} + \varphi\left(\frac{1}{c^2_{\varphi,J}}(y-\mu)^T\Sigma^{-1}(y-\mu)\right)\right] \mathrm{d}F_{(\mu_0,\Sigma_0)}(y) \tag{7}$$

defined for $\mu \in \mathbb{R}^J$ and positive-definite, symmetric $\Sigma \in \mathbb{R}^J_J$, attains the global minimum if and only if $\mu = \mu_0$ and $\Sigma = \Sigma_0$.

Proof. The proof falls naturally into two parts. Firstly the equation (7) can be rewritten as expectation and then it can be noticed that the global minimum is attained for $\mu = \mu_0$. The next step is to replace μ by μ_0 and to note that the global minimum is attained for $\Sigma = \Sigma_0$ (more details can be found in Kulawik and Zontek's article [5]). ∎

Theorem 7 (Fisher consistency)**.** *Assume that* $\theta_0 \in \Theta$*. If* $\varphi\colon [0,+\infty) \to \mathbb{R}$ *satisfies* (B1) *and* (B2)*, then*

$$\theta \longrightarrow \int \Phi(y|\theta)\,\mathrm{d}F_{\theta_0}(y), \quad \theta \in \Theta,$$

with $c = c_{\varphi,J}$*, attains the global minimum if and only if* $\theta = \theta_0$*.*

Proof. If $\theta_0 \in \Theta$, then $\theta_0 = (\beta_0^T, \alpha_{0_1}, \ldots, \alpha_{0_k})^T$ for some $\beta_0 \in \mathbb{R}^a$ and $\alpha_{0_i} \in (0,+\infty)$, $i = 1,\ldots,k$, and $\sum_{i=1}^k \alpha_{0_i} W_i > 0$. Now we can use Lemma 6 with $\mu_0 = X\beta_0$ and $\Sigma_0 = \sum_{i=1}^k \alpha_{0_i} W_i$ to end the proof. ∎

Now we can move on to the part concerning Fréchet differentiability at F_θ, $\theta = (\beta^T, \alpha^T)^T \in \Theta$, of functional T^*. Let us introduce the following notation:

$$\Psi(\cdot|\theta) = \left(\Psi^{(1)}(\cdot|\theta)^T, \Psi^{(2)}(\cdot|\theta)^T\right)^T = \left(\frac{\partial \Phi(\cdot|\theta)}{\partial \beta}^T, \frac{\partial \Phi(\cdot|\theta)}{\partial \alpha}^T\right)^T,$$

$$M(\theta) = \int \left[\frac{\partial \Psi(y|\theta)}{\partial \theta}\right] \mathrm{d}F_\theta(y),$$

$$\Sigma = \sum_{i=1}^k \alpha_i W_i, \quad \tilde{D}_i = \Sigma^{-\frac{1}{2}} W_i \Sigma^{-\frac{1}{2}}, \quad i \in \{1,\ldots,k\},$$

and

$$\tilde{D} = \left[\mathrm{tr}(\tilde{D}_i \tilde{D}_j)\right]_{1\le i,j\le k}, \quad \tilde{d} = \left(\mathrm{tr}(\tilde{D}_1), \ldots, \mathrm{tr}(\tilde{D}_k)\right)^T.$$

Lemma 8. *If* $\varphi\colon [0,+\infty) \to \mathbb{R}$ *satisfies* (B1)-(B4) *and if* ξ *is a chi-squared variable with* J *degrees of freedom, then*

(i) *for* $w_1 = \frac{4}{J c_{\varphi,J}^2}\mathbb{E}\left\{\frac{\xi}{c_{\varphi,J}^2}\left[\varphi'\left(\frac{\xi}{c_{\varphi,J}^2}\right)\right]^2\right\}$ *we have*

$$\int \Psi^{(1)}(y|\theta)\left[\Psi^{(1)}(y|\theta)\right]^T \mathrm{d}F_\theta(y) = w_1 X^T \Sigma^{-1} X$$

and

$$\int \frac{\partial \Psi^{(1)}(y|\theta)}{\partial \beta}\,\mathrm{d}F_\theta(y) = X^T \Sigma^{-1} X,$$

(ii) *for* $w_2 = \frac{8}{J(J+2)}\mathbb{E}\left\{\left[\frac{\xi}{c_{\varphi,J}^2}\varphi'\left(\frac{\xi}{c_{\varphi,J}^2}\right)\right]^2\right\}$ *we have*

$$\int \Psi^{(2)}(y|\theta)\left[\Psi^{(2)}(y|\theta)\right]^T \mathrm{d}F_\theta(y) = \frac{1}{4}\left[w_2 \tilde{D} + \left(\frac{1}{2}w_2 - 1\right)\tilde{d}\tilde{d}^T\right]$$

and for $w_3 = \frac{4}{J(J+2)}\mathbb{E}\left[\left(\frac{\xi}{c_{\varphi,J}}\right)^2 \varphi'\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right]$ *we have*

$$\int \frac{\partial \Psi^{(2)}(y|\theta)}{\partial \alpha}\, \mathrm{d}F_\theta(y) = \frac{1}{4}\left[w_3 \tilde{D} + \left(\frac{1}{2}w_3 - 1\right)\tilde{d}\tilde{d}^T\right],$$

(iii) *we have*

$$\int \Psi^{(1)}(y|\theta)\left[\Psi^{(2)}(y|\theta)\right]^T \mathrm{d}F_\theta(y) = \left[\int \frac{\partial \Psi^{(2)}(y|\theta)}{\partial \beta}\, \mathrm{d}F_\theta(y)\right]^T = 0^a_k.$$

Proof. Let $Z = (Z_1, \ldots, Z_J)^T = \Sigma^{-\frac{1}{2}}(Y - X\beta)$, where Y has the distribution $\mathrm{N}_J(X\beta, \Sigma)$.

(i) The following equations

$$\Psi^{(1)}(y|\theta) = \frac{-2}{c^2_{\varphi,J}} X^T \Sigma^{-\frac{1}{2}} z \varphi'\left(\frac{z^T z}{c^2_{\varphi,J}}\right),$$

$$\frac{\partial \Psi^{(1)}(y|\theta)}{\partial \beta} = \frac{2}{c^2_{\varphi,J}} X^T \Sigma^{-\frac{1}{2}}\left[\frac{2}{c^2_{\varphi,J}} z z^T \varphi''\left(\frac{z^T z}{c^2_{\varphi,J}}\right) + I_J \varphi'\left(\frac{z^T z}{c^2_{\varphi,J}}\right)\right]\Sigma^{-\frac{1}{2}} X$$

and also (12), (11) and (15) are sufficient to end the proof of (i).

(ii) For $i, j \in \{1, \ldots, k\}$ we have

$$\frac{\partial \Phi(y|\theta)}{\partial \alpha_i} = \frac{1}{2}\mathrm{tr}(\tilde{D}_i) - \frac{1}{c^2_{\varphi,J}} z^T \tilde{D}_i z \varphi'\left(\frac{z^T z}{c^2_{\varphi,J}}\right)$$

and

$$\begin{aligned}\frac{\partial\left(\frac{\partial \Phi(y|\theta)}{\partial \alpha_i}\right)}{\partial \alpha_j} &= -\frac{1}{2}\mathrm{tr}(\tilde{D}_i\tilde{D}_j) + \frac{1}{c^4_{\varphi,J}} z^T \tilde{D}_j z z^T \tilde{D}_i z \varphi''\left(\frac{z^T z}{c^2_{\varphi,J}}\right)\\ &\quad + \frac{1}{c^2_{\varphi,J}} z^T \tilde{D}_j \tilde{D}_i z \varphi'\left(\frac{z^T z}{c^2_{\varphi,J}}\right) + \frac{1}{c^2_{\varphi,J}} z^T \tilde{D}_i \tilde{D}_j z \varphi'\left(\frac{z^T z}{c^2_{\varphi,J}}\right).\end{aligned}$$

Part (ii) is now a consequence of (14), (13) and (17).

(iii) A proof is similar — one can use (10) in the proof. ■

Remark 9. A consequence of (16) *is that* $w_1 \geq 1$.

Theorem 10 (Fréchet differentiability). *If* $\varphi\colon [0,+\infty) \to \mathbb{R}$ *satisfies* (B1)–(B4), *then*

$$T^*(G) - T^*(F_\theta) = -M(\theta)^{-1} \int \psi(y|\theta)\, \mathrm{d}G(y) + o\left(||G - F_\theta||_\infty\right).$$

Proof. Fréchet differentiability of T^* at F_θ results from Clarke's [1] conditions which are fulfilled through (B1)–(B4). ■

Let $\hat{F_N}$ stand for the empirical distribution function of a random sample consisting of N vectors, each having identical distribution from infinitesimal neighborhood of F_θ.

Theorem 11. *If* $\varphi\colon [0,+\infty) \to \mathbb{R}$ *satisfies* (B1)-(B4), *then distribution of* $T^*(\hat{F_N})$ *can be approximated with multivariate normal distribution with mean vector* θ *and covariance matrix*

$$\frac{1}{N}\begin{bmatrix} w_1\left(X^T\Sigma^{-1}X\right)^{-1} & 0_k^a \\ 0_a^k & 4\left(\gamma_1\tilde{D}^{-1} + \gamma_2\tilde{D}^{-1}\tilde{d}\tilde{d}^T\tilde{D}^{-1}\right)\end{bmatrix}, \tag{8}$$

where

$$\gamma_1 = \frac{w_2}{w_3^2},\ \ \gamma_2 = \frac{2w_3(4w_2 - 2w_3 - w_2w_3) - w_2(w_3-2)^2\tilde{d}^T\tilde{D}^{-1}\tilde{d}}{\left[2w_3^2 + w_3\left(w_3-2\right)\tilde{d}^T\tilde{D}^{-1}\tilde{d}\right]^2}$$

and w_1, w_2, w_3 *can be found in Lemma* 8.

Proof. The theorem is a consequence of

$$M(\theta) = \begin{bmatrix} X^T\Sigma^{-1}X & 0_k^a \\ 0_a^k & \frac{1}{4}\left[w_3\tilde{D} + \left(\frac{1}{2}w_3 - 1\right)\tilde{d}\tilde{d}^T\right]\end{bmatrix}$$

and

$$\int \Psi(y|\theta)\left[\Psi(y|\theta)\right]^T \mathrm{d}F_\theta(y) = \begin{bmatrix} w_1X^T\Sigma^{-1}X & 0_k^a \\ 0_a^k & \frac{1}{4}\left[w_2\tilde{D} + \left(\frac{1}{2}w_2 - 1\right)\tilde{d}\tilde{d}^T\right]\end{bmatrix}.$$

The matrices above follow from Lemma 8. It is sufficient to use Theorem 3 to end the proof. ■

Remark 12. Under the model assumptions, for m.l.e. we can get matrix of the form (8) with $w_1 = 1$, $\gamma_1 = \frac{1}{2}$, $\gamma_2 = 0$ ($w_2 = w_3 = 2$). Distribution of the estimator can be approximated then with multivariate normal distribution with mean vector θ and covariance matrix

$$\frac{1}{N}\begin{bmatrix}\left(X^T\Sigma^{-1}X\right)^{-1} & 0_k^a \\ 0_a^k & 2\tilde{D}^{-1}\end{bmatrix}. \tag{9}$$

4. Computer simulation results

In simulation studies model of distribution

$$\mathrm{N}_3\left((\mu_1,\mu_2,\mu_3)^T, \sum_{i=1}^{6}\alpha_i W_i\right)$$

with parameters $\mu_i \in \mathbb{R}$, $i = 1,2,3$, $\alpha_j \in \mathbb{R}$, $j = 1,\dots,6$, and matrices

$$W_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; W_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; W_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$W_4 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; W_5 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \; W_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

was considered. The model has the form $\mathrm{N}_J(X\beta, \Sigma)$ with $J = 3$, $X = I_3$ and $\beta = (\mu_1,\mu_2,\mu_3)^T$. Two estimators of parameter $\theta = (\mu_1,\mu_2,\mu_3,\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6)^T$ are compared: m.l.e. and robust estimator described in Section 3 (r.e.). The estimators can be defined by statistical functionals of form (4) with corresponding objective functions. In case of m.l.e. the objective function is given by (5) with $\varphi(x) = \frac{1}{2}x$, $x \geq 0$ and $c = 1$. In case of r.e. the objective function is given by (5) with $\varphi(x) = \phi(\sqrt{x})$, $x \geq 0$, and the function ϕ is defined by its derivative

$$\phi'(x) = \begin{cases} x, & |x| \leq t, \\ -x - 4t - \frac{2t^2}{x}, & -2t < x < -t, \\ -x + 4t - \frac{2t^2}{x}, & t < x < 2t, \\ \frac{2t^2}{x}, & |x| \geq 2t. \end{cases}$$

In simulation $t = 1,577$ was taken to get $w_1 = 1.1$ ($c_{\varphi,J} = 0.847$, $\gamma_1 = 0.574$, $\gamma_2 = 0.040$) in matrix (8). In case of m.l.e. we get the matrix (9).

From the model $\mathrm{N}_3\left(0^3, W\right)$, where $W = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, 20 observation vectors (model vectors) were generated randomly. For the observations estimates of parameter θ has been noted. The procedure was repeated 5000 times. The next simulation was involved after changing 5% of the model data for vectors taking from respectively

(i) $\mathrm{N}_3\left((20,0,0)^T, W\right)$,

(ii) $\mathrm{N}_3\left((20,20,0)^T, W\right)$,

(iii) $\mathrm{N}_3\left((20,20,10)^T, W\right)$,

(iv) $\mathrm{N}_3\left(0^3, \begin{bmatrix} 50 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}\right)$,

(v) $\mathrm{N}_3\left(0^3, \begin{bmatrix} 50 & 25 & 25 \\ 25 & 50 & 25 \\ 25 & 52 & 25 \end{bmatrix}\right)$.

In cases of (i)–(iii) shift parameter of model distribution is contaminated. In the other cases — its covariance matrix.

Results of the computer simulation are given in Table 1 presented below. The table consists of ten columns. In the first one there is a name of estimator (m.l.e., r.e.) and a type of contamination, when it is needed. The other columns — labeled by a name of parameters μ_1, μ_2, μ_3, α_1, α_2, α_3, α_4, α_5 or α_6, present the information concerning their estimation. The results of the case described by the first column of Table 1 are given in three rows. In the first one there is the average of 5000 estimates for each parameter (theoretical value of parameter). The second row presents sample standard deviations (s.s.d.) and the third one - theoretical standard deviations (t.s.d.) that are equal to square root of diagonal elements of matrices (9) and (8) in case of m.l.e. and r.e., respectively.

Results from Table 1 show that contamination of chosen coordinates affects the theoretical values and s.s.d. for corresponding parameters in case of m.l.e. Meanwhile the values concerning the other parameters seem not to change (in comparison with model data case). For example in the case of (i) the first of coordinates of model data is contaminated. Results in the rows labeled by "m.l.e. (i)" shows that theoretical values and s.s.d. for parameters μ_1 and α_1 have significantly increased. For r.e. a contamination concerning even one of the coordinates affects only insignificantly theoretical values and s.s.d. for all parameters. What is more, the bigger number of coordinates is contaminated, the better results of estimation gives r.e. For r.e. all theoretical values of parameters are much closer the model ones than in case of m.l.e. For r.e. all s.s.d. are much closer corresponding t.s.d. than in case of m.l.e. It seems that the approximation of r.e.'s distribution described in Theorem 11 is sensible both for model and little contaminated data.

Table 1

	μ_1	μ_2	μ_3	α_1	α_2	α_3	α_4	α_5	α_6
	0	0.004	0	1.907	1.894	0.951	-0.001	0.952	0.002
m.l.e.	(0.317)	(0.314)	(0.225)	(0.609)	(0.607)	(0.305)	(0.437)	(0.373)	(0.307)
	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
	-0.001	0.004	0	1.939	1.918	0.965	-0.002	0.967	0.002
r.e.	(0.332)	(0.333)	(0.235)	(0.682)	(0.681)	(0.339)	(0.480)	(0.414)	(0.334)
	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)
m.l.e.	0.983	0.004	0	19.658	1.894	0.951	0.013	0.961	0.002
	(1.025)	(0.314)	(0.225)	(16.960)	(0.607)	(0.305)	(1.401)	(1.022)	(0.307)
(i)	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
r.e.	0.049	0.004	0	2.896	1.842	0.925	0.001	0.929	0.003
	(0.449)	(0.340)	(0.241)	(5.709)	(0.671)	(0.335)	(0.505)	(0.441)	(0.330)
(i)	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)
m.l.e.	0.983	0.987	0	19.658	19.671	0.951	17.763	0.961	0.011
	(1.025)	(1.025)	(0.225)	(16.960)	(16.907)	(0.305)	(16.805)	(1.022)	(1.006)
(ii)	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
r.e.	0.039	0.043	0	2.672	2.656	0.925	0.814	0.927	0.004
	(0.441)	(0.440)	(0.241)	(5.357)	(5.419)	(0.334)	(5.354)	(0.435)	(0.365)
(ii)	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)
m.l.e.	0.983	0.987	0.491	19.658	19.671	5.397	17.763	9.836	8.892
	(1.025)	(1.025)	(0.533)	(16.960)	(16.907)	(4.309)	(16.805)	(8.517)	(8.439)
(iii)	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
r.e.	0.048	0.052	0.024	2.869	2.849	1.179	1.007	1.433	0.508
	(0.445)	(0.445)	(0.282)	(5.540)	(5.570)	(1.444)	(5.530)	(2.809)	(2.800)
(iii)	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)
m.l.e.	-0.001	0.004	0	4.205	1.894	0.951	-0.004	0.949	0.002
	(0.465)	(0.314)	(0.225)	(4.172)	(0.607)	(0.305)	(0.647)	(0.503)	(0.307)
(iv)	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
r.e.	0	0.004	0	2.200	1.880	0.946	-0.001	0.946	0.003
	(0.345)	(0.336)	(0.238)	(0.907)	(0.674)	(0.337)	(0.487)	(0.417)	(0.332)
(iv)	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)
m.l.e.	0.001	0.008	0.002	4.167	4.080	2.056	1.134	2.055	1.133
	(0.465)	(0.465)	(0.332)	(4.025)	(3.938)	(1.965)	(2.828)	(2.257)	(2.256)
(v)	(0.316)	(0.316)	(0.224)	(0.632)	(0.632)	(0.316)	(0.447)	(0.387)	(0.316)
r.e.	-0.002	0.003	-0.001	2.104	2.069	1.042	0.082	1.043	0.086
	(0.341)	(0.341)	(0.241)	(0.786)	(0.766)	(0.384)	(0.538)	(0.465)	(0.381)
(v)	(0.332)	(0.332)	(0.235)	(0.701)	(0.701)	(0.350)	(0.479)	(0.425)	(0.339)

References

[1] B.R. Clarke, *Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations*, Ann. Statist. **11** (4) (1983) 1196–1205.

[2] T. Bednarski and S. Zontek, *Robust estimation of parameters in a mixed unbalanced model*, Ann. Statist. **24** (4) (1996) 1493–1510. doi:10.1214/aos/1032298279

[3] P.J. Huber, Robust Statistics (Wiley, New York, 1981). doi:10.1002/0471725250

[4] J. Kiefer, *On large deviations of the empiric D.F. of vector chance variables and a law of iterated logarithm*, Pacific J. Math. **11** (1961) 649–660. doi:10.2140/pjm.1961.11.649

[5] A. Kulawik and S. Zontek, *Robust estimation in the multivariate normal model with variance components*, Statistics **49** (4) 766–780. doi:10.1080/02331888.2014.932793

[6] R.A. Maronna, *Robust M-estimators of multivariate location and scatter*, Ann. Statist. **4** (1) (1976) 51–67. doi:10.1214/aos/1176343347

[7] P.J. Rousseeuw, *Multivariate estimation with high breakdown point*, Mathematical Statistics and Applications, Vol. B (Bad Tatzmannsdorf, 1983), (Reidel, Dordrecht, 1985) 283–297.

[8] R. Zmyślony and S. Zontek, *Robust M-estimator of parameters in variance components model*, Discuss. Math. Probability and Statistics **22** (2002) 61–71.

[9] S. Zontek, *Multivariate robust estimation in linear model for spatially located sensors and random input*, Discuss. Math. Algebra and Stochastic Methods **18** (1998) 195–206.

Received 3 February 2016
Revised 18 July 2016

APPENDIX

The appendix consists of three lemmas that have been used in the proof of Lemma 8. Proofs of them have been shortly explained (more details can be found in Kulawik and Zontek's article [5]).

Lemma 13. *Let $Z = (Z_1, \dots, Z_J)^T$ be a multivariate normally distributed random vector with mean vector 0^J and covariance matrix I_J. Assume that $f\colon [0, +\infty) \to \mathbb{R}$ is such that the expected values below exist.*

(i) *If $n_1, \dots, n_J$ and for some i, n_i is an odd number, then*

$$\mathbb{E}\left(Z_1^{n_1} \cdot \ldots \cdot Z_J^{n_J} f(Z^T Z)\right) = 0. \tag{10}$$

(ii) *If ξ is a chi-squared variable with J degrees of freedom, then*

$$\mathbb{E}\left(Z_1^2 f(Z^T Z)\right) = \frac{1}{J}\mathbb{E}\big(\xi f(\xi)\big) \tag{11}$$

and

$$\mathbb{E}\left(ZZ^T f(Z^T Z)\right) = \mathbb{E}\left(Z_1^2 f(Z^T Z)\right) I_J. \tag{12}$$

If $M_1, M_2 \in \mathbb{R}_J^J$ are symmetric, then

$$\mathbb{E}\big(Z^T M_1 Z Z^T M_2 Z f(Z^T Z)\big) = \frac{1}{J(J+2)}\mathbb{E}\big(\xi^2 f(\xi)\big)\big(2\mathrm{tr}(M_1 M_2) + \mathrm{tr}(M_1)\mathrm{tr}(M_2)\big). \tag{13}$$

Proof. Main idea of the proof is to express the above expected values in polar coordinates. ■

Lemma 14. *Let $Z = (Z_1, \ldots, Z_J)^T$ be a multivariate normally distributed random vector with mean vector 0^J and covariance matrix I_J. If $\varphi\colon [0, +\infty) \to \mathbb{R}$ satisfies* (B1)-(B4) *and if $A \in \mathbb{R}^J_J$, then*

$$\mathbb{E}\left[\frac{1}{c^2_{\varphi,J}} Z^T A Z \varphi'\left(\frac{Z^T Z}{c^2_{\varphi,J}}\right)\right] = \frac{1}{2}\mathrm{tr}(A). \tag{14}$$

Proof. The equation (14) is a consequence of both the equation (12) and the way of choosing the constant $c_{\varphi,J}$. ■

Lemma 15. *If $\varphi\colon [0, +\infty) \to \mathbb{R}$ satisfies* (B1)–(B4) *and if ξ is a chi-squared variable with J degrees of freedom, then*

$$\frac{2}{c^2_{\varphi,J}}\left\{\frac{2}{J}\mathbb{E}\left[\frac{\xi}{c^2_{\varphi,J}}\varphi''\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right] + \mathbb{E}\left[\varphi'\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right]\right\} = 1, \tag{15}$$

$$\frac{4}{J c^2_{\varphi,J}}\mathbb{E}\left\{\frac{\xi}{c^2_{\varphi,J}}\left[\varphi'\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right]^2\right\} \geq 1 \tag{16}$$

and

$$\begin{aligned} &2\left\{\frac{4}{J(J+2)}\mathbb{E}\left[\left(\frac{\xi}{c^2_{\varphi,J}}\right)^2\varphi''\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right] + 1\right\} \\ &= \frac{4}{J(J+2)}\mathbb{E}\left[\left(\frac{\xi}{c_{\varphi,J}}\right)^2\varphi'\left(\frac{\xi}{c^2_{\varphi,J}}\right)\right]. \end{aligned} \tag{17}$$

Proof. The inequality (16) is a consequence of the fact that the matrix $\left[\sqrt{\xi}, \sqrt{\xi}\varphi'\left(c^{-2}_{\varphi,J}\xi\right)\right]^T \left[\sqrt{\xi}, \sqrt{\xi}\varphi'\left(c^{-2}_{\varphi,J}\xi\right)\right]$ is nonnegative-definite.
The proofs of (15) and (17) are based on differential calculus and are based on the following form of ξ, that is: $\xi = Z^T Z$, where $Z = (Z_1, \ldots, Z_J)^T$ is a properly chosen random vector. In case of equality (15) we have $Z_i = Y_i - \mu$, for $i = 1, \ldots, J$, where $(Y_1, \ldots, Y_J)^T$ is a multivariate normally distributed $\mathrm{N}_J\left(\mu 1^J, I_J\right)$ random vector, $\mu \in \mathbb{R}$. In case of equality (17) we have $Z_i = \sigma^{-1} Y_i$, for $i = 1, \ldots, J$, where $(Y_1, \ldots, Y_J)^T$ is a multivariate normally distributed $\mathrm{N}_J\left(0^J, \sigma^2 I_J\right)$ random vector, $\sigma > 0$. ■