Discussiones Mathematicae
Probability and Statistics 22 (2002) 61–71

ROBUST M-ESTIMATOR OF PARAMETERS IN VARIANCE COMPONENTS MODEL

Roman Zmyślony and Stefan Zontek

Institute of Mathematics
University of Zielona Góra
Podgórna 50, 65–246 Zielona Góra, Poland

e-mail: r.zmyslony@im.uz.zgora.pl
e-mail: s.zontek@im.uz.zgora.pl

Abstract

It is shown that a method of robust estimation in a two way crossed classification mixed model, recently proposed by Bednarski and Zontek (1996), can be extended to a more general case of variance components model with commutative a covariance matrices.

Keywords: Robust estimator, maximum likelihood estimator, statistical functional, Fisher consistency, Fréchet differentiability.

2000 Mathematics Subject Classification: 62F35, 62J05.

1. Introduction and notation

In the paper, we assume that a distribution function of a sample we observed is modeled by the following family of normal distributions on $\mathcal{R}^n$

$$\left\{ \mathrm{N}\left(X\beta,\ \sum_{i=1}^{k} \sigma_i^2 V_i \right) :\ \beta \in \mathcal{R}^p, \sigma_1 \geq 0, \ldots, \sigma_{k-1} \geq 0, \sigma_k > 0 \right\},$$

where X is a known matrix, while $V_1, \ldots, V_k$ are known nonnegative definite matrices. There is ample literature concerning a situation when the true distribution of the sample satisfies the model assumptions. There were proposed best unbiased estimators, admissible estimators, maximum likelihood estimators and so on. However, when the actual distribution is even very closed to the model distribution (in the sense of the supremum norm of the cumulative distribution function) a “classical” estimate can be arbitrary far from any model parameter.

There are different approaches to robust analysis in mixed models. One of them relies on the robust estimation of the random effects and naturally leads to high dimensional nonlinear quations. This concept was introduced to the interlaboratory model by Rocke (1991) (see also Iglewicz, 1993). Another approach based on the modified loglikelihood function. Here we can mention papers by Huggins (1993), Bednarski and Zontek (1996), Bednarski and Clarke (1993). The modification is so chosen that the corresponding statistical functional is Fisher consistent and Fréchet differentiable. This implies important robustness properties (see Bednarski, Clarke and Kołkiewicz, 1991, and Bednarski, 1994).

In the paper, we adopt the approach of Bednarski and Zontek (1996) to the following model

$$N\left(X\beta,\ \sum_{i=1}^{k}\sigma_i^2 V_i\right), \tag{1}$$

where X is a known $n \times p$ matrix of rank p, $V_1, \dots, V_k$ are linearly independent known $n \times n$ nonnegative definite matrices, while $\beta \in \mathcal{R}^p$ and $\sigma_1 \geq 0, \dots, \sigma_{k-1} \geq 0, \sigma_k > 0$ are unknown parameters. In addition, we assume that $V_k = I_n$ and that $V_1, \dots, V_k$ commute.

In Section 2, a statistical functional defining an estimator of the vector of fixed effects and scale components is presented. Under the assumptions appearing in Bednarski and Zontek (1996) it is shown that the functional is Fisher consistent and Fréchet differentiable for the supremum norm. This in turn implies that the corresponding estimators are robust and asymptotically normal. Finally, the explicite form of the asymptotic covariance matrix is given.

Let $\mathcal{M}_{n,r}$ denote the space of $n \times r$ real matrices and $\mathcal{S}_n^{\geq}$ the subspace of $\mathcal{M}_{n,n}$ of all symmetric nonnegative definite matrices. The symbol $A \otimes B$ means the Kronecker product of matrices A and B. Throughout the paper $\mathrm{diag}(w)$ stands for a diagonal matrix with the i-th diagonal element equal to the i-th component of a vector w. The range and the transpose of any matrix W is written as $R(W)$ and W^T, respectively.

2. The variance components model

Let $Y_1, \ldots, Y_N$ be a sample from the model (1). Then the vector $(Y_1^T, \ldots, Y_N^T)^T$ has the following distribution

$$N\left((\mathbf{1}_N \otimes X)\beta, I_N \otimes \sum_{i=1}^{k} \sigma_i^2 V_i\right). \tag{2}$$

The distribution function corresponding to (2) will be denoted by $F(\cdot|\theta)$, where $\theta = (\beta^T, \sigma^T)^T$, while $\sigma = (\sigma_1, \ldots, \sigma_k)^T$.

Under the assumption that $V_1, \ldots, V_k$ commute there exist nonzero matrices $Q_1, \ldots, Q_q$ in $\mathcal{S}_n^{\geq}$ such that

$$\operatorname{span}\{Q_1, \ldots, Q_q\} = \operatorname{qspan}\{V_1, \ldots, V_k\}$$

and that

$$Q_i Q_j = \delta_{ij} Q_i, \quad i, j = 1, \ldots, q$$

(see Seely, 1971, for construction see also Zmyślony and Drygas, 1992), where δ_{ij} is the Kronecker's delta. So each matrix V_i, $i = 1, \ldots, k$, can be presented as

$$V_i = \sum_{j=1}^{q} h_{ij} Q_j.$$

Since $Q_1, \ldots, Q_q$ are idempotent and symmetric matrices, they can be expressed as $Q_i = \boldsymbol{P}_i \boldsymbol{P}_i^T$, where $\boldsymbol{P}_i \in \mathcal{M}_{n \times n_i}$ and $\boldsymbol{P}_i^T \boldsymbol{P}_i = \boldsymbol{P}_i^T Q_i \boldsymbol{P}_i = I_{n_i}$ (this also means that $\operatorname{rank}(Q_i) = n_i$). Then one can easily find that

$$\boldsymbol{P}^T \left(\sum_{i=1}^{k} \sigma_i V_i\right) \boldsymbol{P} = \sum_{j=1}^{q} \left(\sum_{i=1}^{k} \sigma_i^2 h_{ij}\right) \boldsymbol{P}^T Q_j \boldsymbol{P},$$

where $\boldsymbol{P} = (\boldsymbol{P}_1, \ldots, \boldsymbol{P}_q)$, is a diagonal matrix with diagonal elements given by

$$\delta_r^2 = \sum_{i=1}^{k} \sigma_i^2 h_{ij}, \quad r = N_{j-1}, \ldots, N_j, \ \ j = 1, \ldots, q, \tag{3}$$

where $N_o = 0$, $N_j = \sum_{i=1}^{j} n_i$, $j = 1, \ldots, q$.

3. Robust M-functional

Under the assumptions imposed on the model (1) the loglikelihood function can be written as

$$\begin{aligned} l(y|\theta) &= \ln\left(|V_\sigma|^{1/2}\right) + 0.5\,(y - X\beta)^T\, V_\sigma^{-1}(y - X\beta) \\ &= \sum_{r=1}^{n}\left[\ln(\delta_r) + 0.5\left(P_r^T\,(y - X\beta)/(c\delta_r)\right)^2\right]. \end{aligned}$$

A statistical functional will be defined via an objective function, which is a modification of the loglikelihood function. This function can be written in the following form

$$\Phi(y|\theta) = \sum_{r=1}^{n}\left[\ln(\delta_r) + \phi\left(P_r^T\,(y - X\beta)/(c\delta_r)\right)\right], \tag{4}$$

where a function $\phi : \mathcal{R}^n \to \mathcal{R}$ is properly chosen, while P_i is the i-th column of the matrix $\boldsymbol{P}$. The function Φ becomes the loglikelihood function when $\phi(x) = x^2/2$ and $c = 1$. For a given ϕ we can frequently choose c to make the functional Fisher consistent.

Definition 1 (the functional). Define the functional T for a given distribution function G on $\mathcal{R}^n$ to be the parameter θ for which

$$\int \Phi(y|\theta)\mathrm{d}G(y) \tag{5}$$

attains the minimum value.

For a given statistical functional T let us define an estimator $\hat{\theta}_N$ of θ by $\hat{\theta}_N = T(\hat{F}_N)$, where $\hat{F}_N$ is the empirical distribution function based on $Y_1, \ldots, Y_N$.

Similarly as in Bednarski and Zontek (1996), we state the assumptions concerning ϕ which imply the Fisher consistency (**A1**, **A2**) and Fréchet differentiability (**A3**, **A4**) of T.

A1 The function ϕ is symmetric about 0, and it has a positive derivative.

A2 The function $x\phi'(x)$ has a nonnegative derivative for $x \geq 0$ and there exists $x_o > 0$ such that $x_o\phi'(x_o) > 1$.

A3 The functions ϕ' and ϕ'' are bounded.

A4 The functions $x\phi'(x)$ and $x^2\phi''(x)$ are bounded.

Remark. For the robust estimation of scale Huber (1981) proposed the function ϕ given by

$$\phi(x) = \begin{cases} x^2/2, & |x| \leq t, \\ t^2(\ln(|x|) - \ln(t^2) + 0.5), & |x| > t. \end{cases}$$

This function after smooth modification in a neibourhood of points $-t$ and t (for details see Bednarski and Zontek, 1996) satisfies the assumptions **A1**–**A4**.

The assumption **A2** implies that there is a unique $c > 0$ satisfying

$$\mathbf{E}[(W/c)\phi'(W/c) - 1] = 0, \tag{6}$$

where W is the standard normal random variable (see Bednarski and Zontek, 1996).

Theorem 1 (Fisher consistency). *Let $\theta_o = ((\beta^o)^T, (\sigma^o)^T)^T$ be a given parameter and let c in* (4) *be defined by* (6)*. If ϕ satisfies* **A1** *and* **A2**, *then*

$$\int \Phi(y|\theta)\mathrm{d}F(y|\theta_0) \tag{7}$$

attains the global minimum if and only if $\theta = \theta_0$.

Proof. As in Bednarski and Zontek (1996), we can reduce the minimization problem to the one dimensional shift and scale case. Really, (7) attains the global minimum when for each $r = 1, \ldots, n$

$$\int \left[\ln(\delta_r) + \phi\left(P_r^T(y - X\beta)/(c\delta_r)\right)\right] \mathrm{d}F(y|\theta)$$

attains the global minimum. So for $j = 1, \ldots, q$

$$\boldsymbol{P}_j^T X\beta = \boldsymbol{P}_j^T X\beta^o \text{ and } \sum_{i=1}^{k} \sigma_i^2 h_{ij} = \sum_{i=1}^{k} (\sigma_i^o)^2 h_{ij}.$$

This means that $\beta = \beta^o$ and $\sigma = \sigma^o$. ■

Below we give a technical lemma which implies a simple form of the asymptotic covariance matrix of the estimator $\hat{\theta}_N$.

Define for $\sigma = (\sigma_1, \ldots, \sigma_k)^T$ the following matrices

$$U^{(1)}(\sigma) = X^T V_\sigma^{-1} X$$

and

$$U^{(2)}(\sigma) = 2\sum_{j=1}^{q} \frac{n_j}{\left(\sum_{i=1}^{k} \sigma_i^2 h_{ij}\right)^2} \operatorname{diag}(\sigma) H_j H_j^T \operatorname{diag}(\sigma),$$

where $V_\sigma = \sum_{i=1}^{k} \sigma_i^2 V_i$, while $H_j = (h_{1j}, \ldots, h_{kj})^T$.

Let $\Psi^{(1)}(\cdot|\theta)$ and $\Psi^{(2)}(\cdot|\theta)$ be a partition of the vector function $\Psi(\cdot|\theta) = \frac{\partial}{\partial\theta}\Phi(\cdot|\theta)$ with respect to fixed effects and scale components, respectively. Moreover, let

$$\Delta(\cdot|\theta) = \begin{bmatrix} \Delta^{11}(\cdot|\theta) & \Delta^{12}(\cdot|\theta) \\ \Delta^{12}(\cdot|\theta)^T & \Delta^{22}(\cdot|\theta) \end{bmatrix}$$

be the corresponding partition of the matrix $\frac{\partial}{\partial\theta}\Psi(\cdot|\theta)$.

Lemma 1. *Let the constant c be defined by* (6) *and let W be a standard normal random variable. If* **A1**, **A2** *are satisfied, then we have*

$$\text{(i)} \quad \int \Psi^{(1)}(y|\theta)\left[\Psi^{(1)}(y|\theta)\right]^T \mathrm{d}F(y|\theta) = c^{-2}\mathbf{E}\left[\phi'(W/c)^2\right] U^{(1)}(\sigma),$$

$$\int \Delta^{11}(y|\theta)\mathrm{d}F(y|\theta) = c^{-2}\mathbf{E}\left[\phi''(W/c)\right] U^{(1)}(\sigma),$$

(ii) $$\int \Psi^{(2)}(y|\theta)\left[\Psi^{(2)}(y|\theta)\right]^T \mathrm{d}F(y|\theta) = 0.5\mathbf{E}\left\{[(W/c)\phi'(W/c)-1]^2\right\}U^{(2)}(\sigma),$$

$$\int \Delta^{22}(y|\theta)\mathrm{d}F(y|\theta) = 0.5\mathbf{E}\left[(W/c)^2\phi''(W/c)+1\right]U^{(2)}(\sigma),$$

(iii) $$\int \Psi^{(1)}(y|\theta)\left[\Psi^{(2)}(y|\theta)\right]^T \mathrm{d}F(y|\theta) = \int \Delta^{12}(y|\theta)\mathrm{d}F(y|\theta) = 0.$$

Proof. Let $\theta = (\beta^T, \sigma^T)^T$ be an arbitrary parameter. Define a random vector

$$Z = (Z_1, \ldots, Z_n)^T = \mathrm{diag}(\delta_1, \ldots, \delta_n)^{-1}\boldsymbol{P}^T(Y - X\beta),$$

where the distribution function of a random vector Y is $F(\cdot|\theta)$. Then components of Z are independent standard normal variables.

Part (i). It is easy to see that the random vector $\Psi^{(1)}(Y|\theta)$ can be represented as

$$\Psi^{(1)}(Y|\theta) = -\sum_{r=1}^{n}\frac{\phi'(Z_r/c)}{c\delta_r}\, X^T P_r.$$

Since

$$\mathbf{E}[\phi'(Z_i/c)\phi'(Z_j/c)] = \begin{cases} \mathbf{E}[\phi'(W/c)^2], & i = j \\ 0, & i \neq j \end{cases},$$

we have

$$\begin{aligned}\mathbf{E}\left\{\Psi^{(1)}(Y|\theta)[\Psi^{(1)}(Y|\theta)]^T\right\} &= \frac{\mathbf{E}[\phi'(W/c)^2]}{c^2}X^T\left(\sum_{r=1}^{n}\frac{1}{\delta_r^2}P_rP_r^T\right)X \\ &= \frac{\mathbf{E}[\phi'(W/c)^2]}{c^2}X^T V_\sigma^{-1}X.\end{aligned}$$

The second equation in part (i) of the lemma can be shown to hold in a similar way, by using the relation

$$\Delta^{11}(Y|\theta) = \sum_{r=1}^{n} \frac{\phi''(Z_r/c)}{c^2\delta_r^2} X^T P_r P_r^T X.$$

Part (ii). The first and the second partial derivatives of $\Phi(Y|\theta)$ with respect to σ are given by

$$\Psi^{(2)}(Y|\theta) = -\sum_{r=1}^{n} \left[\phi'(Z_r/c)Z_r/c - 1\right] \frac{1}{\delta_r} \left(\frac{\partial}{\partial\sigma}\delta_r\right),$$

$$\Delta^{(22)}(Y|\theta) = \sum_{r=1}^{n} \Bigg\{ \left[\phi''(Z_r/c)Z_r^2/c^2 + 1\right] \frac{1}{\delta_r^2} \left(\frac{\partial}{\partial\sigma}\delta_r\right)\left(\frac{\partial}{\partial\sigma}\delta_r\right)^T$$

$$+ \left[\phi'(Z_r/c)Z_r/c - 1\right] \left[\frac{1}{\delta_r}\left(\frac{\partial^2}{\partial^2\sigma}\delta_r\right) - \frac{2}{\delta_r^2}\left(\frac{\partial}{\partial\sigma}\delta_r\right)\left(\frac{\partial}{\partial\sigma}\delta_r\right)^T\right] \Bigg\}.$$

Hence

$$\int \Psi^{(2)}(y|\theta)[\Psi^{(2)}(y|\theta)]^T \mathrm{d}F(y|\theta) = e\sum_{r=1}^{n} \frac{1}{\delta_r^2}\left(\frac{\partial}{\partial\sigma}\delta_r\right)\left(\frac{\partial}{\partial\sigma}\delta_r\right)^T$$

$$= e\sum_{j=1}^{q} \frac{n_j}{\tilde{q}_j}\left(\frac{\partial}{\partial\sigma}\sqrt{\sum_{i=1}^{k}\sigma_i^2 h_{ij}}\right)\left(\frac{\partial}{\partial\sigma}\sqrt{\sum_{i=1}^{k}\sigma_i^2 h_{ij}}\right)^T$$

$$= e\sum_{j=1}^{q} \frac{n_j}{\tilde{q}_j^2}\,\mathrm{diag}(\sigma)H_j H_j^T\,\mathrm{diag}(\sigma),$$

where $e = \mathbf{E}\{[(W/c)\phi'(W/c) - 1]^2\}$, while $\tilde{q}_j = \sum_{i=1}^{k}\sigma_i^2 h_{ij}$. Now making use of equation (6) defining the constant c one can easily derive the second formula of (ii).

Part (iii) follows similarly. ∎

If the function ϕ satisfies the assumptions **A3** and **A4**, then the vector function Ψ fulfills the condition A of Clarke (1983, see also 1996). Therefore the resulting functional T is Fréchet differentiable for the supremum norm (to be denoted by $||\cdot||$). We phrase it in the following theorem.

Theorem 2 (Fréchet differentiability). *If the function ϕ associated with the statistical functional T satisfies* **A3** *and* **A4***, then*

$$T(G) - T(F(\cdot|\theta)) = M(\theta)^{-1} \int \Psi(y|\theta)\mathrm{d}G(y) + o(\|G - F(\cdot|\theta)\|),$$

where

$$M(\theta) = \int \Delta(y|\theta)\mathrm{d}F(y|\theta).$$

Theorems 1, 2 and Kiefer's inequality (Kiefer, 1961) imply that if $\sqrt{N}\|G_N - F(\cdot|\theta)\|$ stays bounded, then

$$\sqrt{N}(\hat{\theta}_N - \theta) = \frac{1}{\sqrt{N}} M(\theta)^{-1} \sum_{i=1}^{N} \Psi(Y_i|\theta) + o_{G_N^{\otimes N}}(1).$$

Since $\Psi(Y_1|\theta), \ldots, \Psi(Y_N|\theta)$ are i.i.d. random vectors with a finite second moment, the central limit theorem implies that $\sqrt{N}(\hat{\theta}_N - \theta)$ is asymptotically normal with zero mean (at the model) and with covariance matrix

$$V(\theta) = M(\theta)^{-1} \left\{ \int \Psi(y|\theta)[\Psi(y|\theta)]^T \, \mathrm{d}F(y|\theta) \right\} M(\theta)^{-1}. \tag{8}$$

Now from Lemma 1 it follows that the covariance matrix for the Fréchet differentiable estimator generated by (5) is of the form

$$V(\theta) = \begin{bmatrix} w_1[U^{(1)}(\sigma)]^{-1} & 0 \\ 0 & w_2[U^{(2)}(\sigma)]^{-1} \end{bmatrix}, \tag{9}$$

where the positive constants w_1 and w_2 are given by the formulas

$$w_1 = \frac{c^2 \mathbf{E}[\phi'(W/c)^2]}{[\mathbf{E}\phi''(W/c)]^2},$$

$$w_2 = \frac{2\mathbf{E}\{[(W/c)\phi'(W/c) - 1]^2\}}{\{\mathbf{E}[(W/c)^2\phi''(W/c) + 1]\}^2},$$

while W is a standard normal random variable.

Note that (9) with $w_1 = w_2 = 1$ is the asymptotic covariance matrix of the maximum likelihood estimator (only at the model). Thus coefficients w_1 and w_2 (always $w_1, w_2 \geq 1$) can be interpreted as asymptotic efficiency coefficients.

Formula (8) gives a possibility of approximating the covariance matrix of the estimator by the following matrix

$$N\hat{M}(\hat{\theta}_N)^{-1}\left\{\int \Psi(\cdot|\hat{\theta}_N)[\Psi(\cdot|\hat{\theta}_N)]^T \mathrm{d}\hat{F}_N\right\}\hat{M}(\hat{\theta}_N)^{-1}, \tag{10}$$

where

$$\hat{M}(\hat{\theta}_N) = \int \Delta(\cdot|\hat{\theta}_N) d\hat{F}_N.$$

References

[1] T. Bednarski, *Fréchet differentiability and robust estimation*, Asymptotic Statistics. Proc. of the Fifth Prague Symp. Physica Verlag, Springer, (1994), 49–58.

[2] T. Bednarski, B.R. Clarke and W. Kołkiewicz, *Statistical expansions and locally uniform Fréchet differentiability*, J. Australian Math. Soc., Ser. **A 50** (1991), 88–97.

[3] T. Bednarski and B.R. Clarke, *Trimmed likelihood estimation of location and scale of the normal distribution*, Australian J. Statists. **35** (1993), 141–153.

[4] T. Bednarski and Z. Zontek, *Robust estimation of parameters in mixed unbalanced models*, Ann. Statist. **24** (4) (1996), 1493–1510.

[5] B.R. Clarke, *Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations*, Ann. Statist. **11** (1983), 1196–1206.

[6] B.R. Clarke, *Nonsmooth analysis and Fréchet differentiability of M-functionals*, Probab. Th. Rel. Fields **73** (1986), 197–209.

[7] B. Iglewicz, *Robust scale estimators and confidence intervals for location*, D.C. Hoagling, F. Mosteller and J.W. Tukey, Eds., Understanding Robust and Exploratory Data Analysis, Wiley, New York, (1983), 404–431.

[8] 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.

[9] D.M. Rocke, *Robustness and balance in the mixed model*, Biometrics **47** (1991), 303–309.

[10] J. Seely, *Quadratic subspaces and completness*, Ann. Math. Statist. **42** (1971), 710–721.

[11] R. Zmyślony and H. Drygas, *Jordan algebras and Bayesian quadratic estimation of variance components*, Linear Algebra and its Applications **168** (1992), 259–275.

Received 26 November 2002