ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIII, 4 (1973)

BOGUSŁAWA BEDNAREK-KOZEK (Wrocław)

ON ESTIMATION IN THE MULTIDIMENSIONAL GAUSSIAN MODEL

1. Introduction. Recently, many papers dealing with admissibility of estimators of parametric functions with matrix loss have been published (see, for example, [1]-[3], [5], [6] and [9]).

Kagan and Šalaevskiĭ gave in [5] some characterization of the normal distribution. Considering the linear Gaussian model, they proved that least squares estimators are admissible in the class of unbiased estimators with matrix loss if and only if the distribution of a sample is normal.

The aim of this paper is to show that in the multidimensional Gaussian model an analogical characterization can be given. Moreover, it is shown that in the case of the p-dimensional normal distribution least squares estimators are admissible with matrix loss when $p \leqslant 2$ and can be inadmissible when $p \geqslant 3$. If a covariance matrix Σ is known and $p \geqslant 3$, then least squares estimators are inadmissible.

2. Characterization of the normal distribution by some properties of least squares estimators. Let

$$X = A\Theta + e, \tag{1}$$

where $A = (a_{ij})$ is a known matrix of order $n \times m$, $\Theta = (\Theta_{ij})$ is a unknown parameter matrix of order $m \times p$, and $e = (e_{ij})$ is a random matrix of order $n \times p$.

We assume that

I. $\operatorname{rank} A = m$, $m \leqslant n$, and $\sum_{j=1}^{m} a_{ij}^2 > 0$ for $i = 1, \ldots, n$.

II. $(\Theta_{11}, \ldots, \Theta_{mp}) \in \Omega$, where Ω is an open subset of the $(m \times p)$-dimensional Euclidean space $R^{m\times p}$.

III. $e^i = (e_{i1}, \ldots, e_{ip})$, $i = 1, \ldots, n$, are independent and identically distributed random vectors with distribution function $F(x_1, \ldots, x_p)$ such that $\mathrm{E}\, e^i = 0$ and $\operatorname{Var} e^i = \Sigma$.

In this paper we deal with the problem of estimation of $\boldsymbol{G\Theta}$, where $\boldsymbol{G}$ is a matrix of order $k\times m$, and we postulate that the loss function is a matrix defined by

$$L(\boldsymbol{\eta}, \boldsymbol{G\Theta}) = (\boldsymbol{\eta}-\boldsymbol{G\Theta})(\boldsymbol{\eta}-\boldsymbol{G\Theta})^T, \tag{2}$$

where $\boldsymbol{\eta}$ is a decision matrix of order $k\times p$.

Thus an estimator $\eta_1(\boldsymbol{X})$ is said to be *better than* $\eta_2(\boldsymbol{X})$ with matrix loss given by (2) if, for every $\boldsymbol{\Theta}\in\Omega$,

$$B_{\boldsymbol{\Theta}} = \mathrm{E}_{\boldsymbol{\Theta}}L\big(\eta_2(\boldsymbol{X}), \boldsymbol{G\Theta}\big)-\mathrm{E}_{\boldsymbol{\Theta}}L\big(\eta_1(\boldsymbol{X}), \boldsymbol{G\Theta}\big)$$

is a non-negative definite matrix (for short, $B_{\boldsymbol{\Theta}}\geqslant 0$), and $B_{\boldsymbol{\Theta}_0}\neq 0$ for at least one $\boldsymbol{\Theta}_0\in\Omega$.

An estimator $\eta(\boldsymbol{X})$ is said to be *admissible* if and only if there does not exist an estimator better than $\eta(\boldsymbol{X})$.

THEOREM 1. 1° *If $n\geqslant 2m+1$ and if there exists a non-degenerate matrix $\boldsymbol{G}$ of order $m\times m$ such that $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$ is admissible in the class of unbiased estimators of $\boldsymbol{G\Theta}$, then $F(x_1, \ldots, x_p)$ is a p-dimensional normal (possibly degenerate) distribution.*

2° *If $F(x_1, \ldots, x_p)$ is a p-dimensional non-degenerate normal distribution and if $n\geqslant m$, then, for every $(k\times m)$-matrix $\boldsymbol{G}$ of rank $k\leqslant m$, the estimator $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$ is admissible in the class of unbiased estimators of $\boldsymbol{G\Theta}$.*

Remark. It is easy to prove that Theorem 1 holds when the loss function is quadratic, i.e.

$$L(\boldsymbol{\eta}, \boldsymbol{G\Theta}) = \sum_{i=1}^{k}\sum_{j=1}^{p}\big(\eta_{ij}-(\boldsymbol{G\Theta})_{ij}\big)^2.$$

For $p=1$ this theorem was proved by Kagan and Šalaevskiĭ [5]. To prove Theorem 1 we need two lemmas.

LEMMA 1. *Let*

$$\boldsymbol{Y} = \boldsymbol{X}-\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$$

and

$$\hat{G} = \tilde{G}-\mathrm{E}_0(\tilde{G}|\boldsymbol{Y}),$$

where

$$\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}.$$

Then

(i) *$\hat{G}$ is an unbiased estimator of $\boldsymbol{G\Theta}$, i.e. $\mathrm{E}_{\boldsymbol{\Theta}}\hat{G} = \boldsymbol{G\Theta}$;*

(ii) $\mathrm{E}_{\boldsymbol{\Theta}}(\hat{G}-\boldsymbol{G\Theta})(\hat{G}-\boldsymbol{G\Theta})^T\leqslant \mathrm{E}(\tilde{G}-\boldsymbol{G\Theta})(\tilde{G}-\boldsymbol{G\Theta})^T$ *with equality holding if and only if* $\mathrm{E}_0(G|\boldsymbol{Y}) = 0$.

Proof. Representing $\boldsymbol{Y}$ in the form of

$$\boldsymbol{Y} = \boldsymbol{X}-\boldsymbol{A\Theta}-\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T(\boldsymbol{X}-\boldsymbol{A\Theta}),$$

we see that the distribution of $\boldsymbol{Y}$ does not depend on $\boldsymbol{\Theta}$.

Consequently,

$$\mathrm{E}_{\boldsymbol{\Theta}}\big(\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\big) = \mathrm{E}_0[\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})] = \mathrm{E}_0\tilde{G} = 0.$$

This implies that

$$\mathrm{E}_{\boldsymbol{\Theta}}\hat{G} = \mathrm{E}_{\boldsymbol{\Theta}}\tilde{G} = \boldsymbol{G\Theta}.$$

In order to prove (ii) we observe that

$$(3)\qquad \mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\boldsymbol{G\Theta})(\tilde{G}-\boldsymbol{G\Theta})^T$$
$$= \mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\hat{G})(\tilde{G}-\hat{G})^T+\mathrm{E}_{\boldsymbol{\Theta}}(\hat{G}-\boldsymbol{G\Theta})(\hat{G}-\boldsymbol{G\Theta})^T+\mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\hat{G})(\hat{G}-\boldsymbol{G\Theta})^T+$$
$$+\mathrm{E}_{\boldsymbol{\Theta}}(\hat{G}-\boldsymbol{G\Theta})(\tilde{G}-\hat{G})^T.$$

Since

$$\mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\hat{G})(\hat{G}-\boldsymbol{G\Theta})^T$$
$$= \mathrm{E}_{\boldsymbol{\Theta}}\{\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\,[\boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T(\boldsymbol{X}-\boldsymbol{A\Theta})-\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})]^T\},$$

and since the distributions of $\boldsymbol{Y}$ and $\boldsymbol{X}-\boldsymbol{A\Theta}$ are independent upon $\boldsymbol{\Theta}$, we have

$$\mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\hat{G})(\hat{G}-\boldsymbol{G\Theta})^T = \mathrm{E}_0\{\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\,[\boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}-\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})]^T\}$$
$$= \mathrm{E}_0\{\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\,[\tilde{G}-\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})]^T\} = 0.$$

Using (3), we obtain

$$\mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\boldsymbol{G\Theta})(\tilde{G}-\boldsymbol{G\Theta})^T = \mathrm{E}_{\boldsymbol{\Theta}}(\hat{G}-\boldsymbol{G\Theta})(\hat{G}-\boldsymbol{G\Theta})^T+\mathrm{E}_{\boldsymbol{\Theta}}\big(\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})^T\big).$$

Because the last term on the right-hand side is a non-negative definite matrix, it follows that

$$\mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G}-\boldsymbol{G\Theta})(\tilde{G}-\boldsymbol{G\Theta})^T \geqslant \mathrm{E}_{\boldsymbol{\Theta}}(\hat{G}-\boldsymbol{G\Theta})(\hat{G}-\boldsymbol{G\Theta})^T.$$

To complete the proof, it is sufficient to note that

$$\mathrm{E}_{\boldsymbol{\Theta}}\big(\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})\mathrm{E}_0(\tilde{G}|\boldsymbol{Y})^T\big) = 0$$

if and only if $\mathrm{E}_0(\tilde{G}|\boldsymbol{Y}) = 0$.

LEMMA 2. *Let $\tilde{G}$ and $\boldsymbol{Y}$ be defined as in Lemma* 1. *If $n \geqslant 2m+1$ and if $\mathrm{E}_0(\tilde{G}|\boldsymbol{Y}) = 0$, while $\boldsymbol{G}$ is a non-degenerate $(m\times m)$-matrix, then $F(x_1, \ldots, x_p)$ is a p-dimensional normal (possibly degenerate) distribution.*

Proof. Since $E_0(\tilde{G}|\boldsymbol{Y}) = 0$, we have

$$E_0\big(\boldsymbol{A}^T\boldsymbol{X}\exp(i\operatorname{tr}\boldsymbol{Y}^T\boldsymbol{T})\big) = 0,$$

where $\boldsymbol{T}$ is an arbitrary $(n\times p)$-matrix. If, in addition, $\boldsymbol{T}$ satisfies the condition $\boldsymbol{A}^T\boldsymbol{T} = 0$, then

$$\boldsymbol{Y}^T\boldsymbol{T} = \boldsymbol{X}^T\boldsymbol{T},$$

and, consequently,

$$\boldsymbol{A}^T E_0\big(\boldsymbol{X}\exp(i\operatorname{tr}\boldsymbol{X}^T\boldsymbol{T})\big) = 0,$$

or, equivalently,

$$\sum_{k=1}^{n}\Big[a_{ks}\prod_{\substack{l=1\\ l\neq k}}^{n}\varphi(t_{l1},\ldots,t_{lp})\frac{\partial}{\partial t_{kj}}\varphi(t_{k1},\ldots,t_{kp})\Big] = 0 \tag{4}$$

for every $s = 1,\ldots,m$, $j = 1,\ldots,p$ and t_{uv} $(u = 1,\ldots,n;\ v = 1,\ldots,p)$ such that $\boldsymbol{A}^T\boldsymbol{T} = 0$, where $\boldsymbol{T} = (t_{uv})$, while $\varphi(t_1,\ldots,t_p)$ is the characteristic function of $F(x_1,\ldots,x_p)$.

Because there exists a neighbourhood $\mathscr{U}$ of zero such that $\varphi(t_1,\ldots,t_p) \neq 0$ for $(t_1,\ldots,t_p)\in\mathscr{U}$, equations (4) can be written for $(t_{k1},\ldots,t_{kp})\in\mathscr{U}$, $k = 1,\ldots,m$, such that $\boldsymbol{A}^T\boldsymbol{T} = 0$, where $\boldsymbol{T} = (t_{uv})$, in the form

$$\sum_{k=1}^{n}\Big[a_{ks}\Big(\frac{\partial}{\partial t_{kj}}\varphi(t_{k1},\ldots,t_{kp})\Big)\Big/\varphi(t_{k1},\ldots,t_{kp})\Big] = 0, \tag{5}$$

$$s = 1,\ldots,m,\ j = 1,\ldots,p.$$

Introducing the notation

$$\psi_j(t_1,\ldots,t_p) = \frac{\partial}{\partial t_j}\ln\varphi(t_1,\ldots,t_p),\quad j = 1,\ldots,p, \tag{6}$$

and

$$\boldsymbol{\Psi}(\boldsymbol{V}) = \big(\boldsymbol{\Psi}_{ij}(\boldsymbol{V})\big) = \big(\psi_j(V_{i1},\ldots,V_{ip})\big),$$

$$i = 1,\ldots,k,\ j = 1,\ldots,p,$$

where $\boldsymbol{V} = (V_{ij})$ is a $(k\times p)$-matrix, we can write formula (5), for $\boldsymbol{T}$ satisfying the condition $\boldsymbol{A}^T\boldsymbol{T} = 0$, as follows:

$$\boldsymbol{A}^T\boldsymbol{\Psi}(\boldsymbol{T}) = 0. \tag{7}$$

Because $n > m$ and $\operatorname{rank}\boldsymbol{A} = m$, there exists an $[n\times(n-m)]$-matrix $\boldsymbol{M}$ having the following two properties: the identity matrix can be obtained from $\boldsymbol{M}$ through crossing out m rows, and

$$\boldsymbol{A}^T\boldsymbol{M} = 0. \tag{8}$$

Then, for any matrix $\boldsymbol{S}$ of order $(n-m)\times p$, there holds $\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{S} = 0$, and, consequently, $\boldsymbol{A}^T\boldsymbol{\Psi}(\boldsymbol{M}\boldsymbol{S}) = 0$.

Now, we shall show that the equation in $\boldsymbol{R}$, $\boldsymbol{M}\boldsymbol{R} = \boldsymbol{\Psi}(\boldsymbol{M}\boldsymbol{S})$, has a solution $\boldsymbol{R} = \boldsymbol{\Psi}(\boldsymbol{S})$.

Let $W_{i_j}(\boldsymbol{M}) = W_j(\boldsymbol{I})$ for $j = 1, \ldots, m$, where $W_s(\boldsymbol{M})$ denotes the s-th row of the matrix $\boldsymbol{M}$. Then $W_k(\boldsymbol{R}) = W_{i_k}(\boldsymbol{M}\boldsymbol{R}) = \boldsymbol{\Psi}\big(\boldsymbol{W}_{i_k}(\boldsymbol{M}\boldsymbol{S})\big) = \boldsymbol{\Psi}\big(W_k(\boldsymbol{S})\big) = W_k\big(\boldsymbol{\Psi}(\boldsymbol{S})\big)$ for every k.

Hence $\boldsymbol{R} = \boldsymbol{\Psi}(\boldsymbol{S})$.

In view of this fact, for every $[(n-m)\times p]$-matrix $\boldsymbol{S}$, there holds

$$\boldsymbol{\Psi}(\boldsymbol{M}\boldsymbol{S}) = \boldsymbol{M}\boldsymbol{\Psi}(\boldsymbol{S}). \tag{9}$$

Next, we show that each column of $\boldsymbol{M}$ having the two mentioned properties has at least two elements that are different from zero. One of them, in the j-th column, is $m_{i_j j} = 1$. Now suppose, for the contrary, that there exists j_0 such that $m_{kj_0} = 0$ for all $k \neq i_{j_0}$. Then (8) would imply that $a_{i_{j_0} l} = 0$ for all $l = 1, \ldots, m$ which would contradict assumption I. Thus every column of $\boldsymbol{M}$ has at least two elements different from zero.

Now let $\boldsymbol{M}_1$ be the $[m\times(n-m)]$-matrix obtained from $\boldsymbol{M}$ by omitting the rows with indices $i_1, \ldots, i_m$. Clearly, in every column of $\boldsymbol{M}_1$ there exists at least one element different from zero. Since $n \geqslant 2m+1$, at least one row of $\boldsymbol{M}_1$, say the j-th row, has two elements different from zero, say $m_{j\alpha}$ and $m_{j\beta}$.

Let $\boldsymbol{S}$ stand for a matrix such that all its rows, except the rows with indices α and β, have all elements equal to zero. The α-th and β-th row can have arbitrary elements. Then, in view of (9) and the fact that $\boldsymbol{\Psi}(0, \ldots, 0) = 0$, we obtain

$$\begin{aligned}\boldsymbol{\Psi}\big(m_{j\alpha}(S_{\alpha 1}, \ldots, S_{\alpha p}) + m_{j\beta}(S_{\beta 1}, \ldots, S_{\beta p})\big) \\ = \boldsymbol{\Psi}\big(m_{j\alpha}(S_{\alpha 1}, \ldots, S_{\alpha p})\big) + \boldsymbol{\Psi}\big(m_{j\beta}(S_{\beta 1}, \ldots, S_{\beta p})\big),\end{aligned} \tag{10}$$

where $(S_{i1}, \ldots, S_{ip})$ is the i-th row of $\boldsymbol{S}$.

Since, for $j = 1, \ldots, p$, $S_{\alpha j}$ and $S_{\beta j}$ are arbitrary numbers, it follows from (10) that, for all vectors $\boldsymbol{U}, \boldsymbol{V} \in \mathscr{U}$, there holds

$$\boldsymbol{\Psi}(\boldsymbol{U}+\boldsymbol{V}) = \boldsymbol{\Psi}(\boldsymbol{U}) + \boldsymbol{\Psi}(\boldsymbol{V}).$$

This implies that, for all $(t_1, \ldots, t_p) \in \mathscr{U}$, there is

$$\psi_i(t_1, \ldots, t_p) = \sum_{j=1}^{p} b_{ij} t_j, \quad i = 1, \ldots, p,$$

where $b_{ij} = \psi_i$ (j-th row of $\boldsymbol{I}_p$), $\boldsymbol{I}_p$ being the identity matrix of order $p \times p$.

Consequently, it follows from (6) that, for all $(t_1, \ldots, t_p) \in \mathscr{U}$,

$$\varphi(t_1, \ldots, t_p) = \exp\Big(\sum_{i>j=1}^{p} b_{ij} t_i t_j + \frac{1}{2}\sum_{i=1}^{p} b_{ii} t_i^2\Big). \tag{11}$$

Since φ is a continuous function, we conclude that formula (11) holds for all $t \in R^p$.

Let $\varphi(t_1, \ldots, t_p)$ be the characteristic function of the distribution of the random vector $(X_1, \ldots, X_p)$. Applying Marcinkiewicz's theorem [7] to every linear function of $X_1, \ldots, X_p$, we conclude that $\varphi(t_1, \ldots, t_p)$ is the characteristic function of a p-dimensional normal (possibly degenerate) distribution. Thus, the proof of Lemma 2 is completed.

Proof of Theorem 1. Lemmas 1 and 2 imply immediately part 1°. To prove part 2°, note that under assumptions I, II and III the statistic $\boldsymbol{A}^T\boldsymbol{X}$ is sufficient and complete. This implies then that $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$ is the unique unbiased estimator of $\boldsymbol{G\Theta}$ based on the sufficient statistic $\boldsymbol{A}^T\boldsymbol{X}$. On the other hand, the Rao-Blackwell theorem [8] implies that $\tilde{G}$ is admissible in the class of unbiased estimators of $\boldsymbol{G\Theta}$.

3. On the admissibility of the least squares estimators.

THEOREM 2. *Let $p \leqslant 2$ and let $\boldsymbol{X}$ be a random matrix given by* (1); *let $F(x_1, \ldots, x_p)$ be the distribution function of a p-dimensional non-degenerate normal distribution, and let $\boldsymbol{G}$ be a non-degenerate $(m \times m)$-matrix. Then, under assumptions* I, II *and* III *formulated in Section* 2, $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$ *is an admissible estimator of $\boldsymbol{G\Theta}$ with matrix loss defined by* (2).

To prove Theorem 2 we need the following lemma:

LEMMA 3. *Let $\boldsymbol{Z} = (\boldsymbol{A}^T\boldsymbol{A})^{-1/2}\boldsymbol{A}^T\boldsymbol{X}$ and $\bar{\boldsymbol{\Theta}} = (\boldsymbol{A}^T\boldsymbol{A})^{1/2}\boldsymbol{\Theta}$. Moreover, let Z^i and $\bar{\Theta}^i$ stand for the i-th rows of $\boldsymbol{Z}$ and $\bar{\boldsymbol{\Theta}}$, respectively.*

If $p \leqslant 2$, then Z^i $(i = 1, \ldots, m)$ is an admissible estimator of $\bar{\Theta}^i$ with quadratic loss.

In the proof of Lemma 3 we use the theorem of Stein (cf. [10] and [11]).

Let $\mathscr{X}$ be a finite-dimensional real coordinate space, $\mathscr{Y}$ an arbitrary space and let the σ-algebra $\mathscr{B}$ be a product, $\mathscr{B} = \mathscr{B}_1 \times \mathscr{B}_2$, where $\mathscr{B}_1$ consists of the Borel sets in $\mathscr{X}$, and $\mathscr{B}_2$ is an arbitrary σ-algebra of subsets of $\mathscr{Y}$. Let the measure λ on $\mathscr{X} \times \mathscr{Y}$ be the product $\lambda = \mu\nu$, where μ is a Lebesgue measure on $\mathscr{B}_1$ and ν is an arbitrary probability measure on $\mathscr{B}_2$. Further, let the parameter space $\bar{\Omega}$ coincide with $\mathscr{X}$.

We observe (X, Y) whose distribution for given $\omega \in \bar{\Omega}$ is such that Y is distributed according to ν, and the conditional density of $X - \omega$ given Y

is $p(\cdot|Y)$, a known density. Moreover, we assume that, for all $y \in \mathcal{Y}$,

$$\int_{\mathcal{X}} p(x, y)\,dx = 1 \quad \text{and} \quad \int_{\mathcal{X}} xp(x, y)\,dx = 0.$$

THEOREM (Stein). (a) *If* $\dim \mathcal{X} = 1$ *and, in addition, the condition*

$$\int_{\mathcal{Y}} \Big(\int_{\mathcal{X}} x^2 p(x, y)\,dx \Big)^{3/2} d\nu(y) < \infty$$

is satisfied, then x is an admissible estimator of ω with quadratic loss.

(b) *If* $\dim \mathcal{X} = 2$ *and if, for some* $\delta > 0$, *the condition*

$$\int_{\mathcal{Y}} \Big(\int_{\mathcal{X}} \|x\|^2 \log^{1+\delta} \|x\|^2 p(x, y)\,dx \Big)^2 d\nu(y) < \infty$$

is satisfied, then x is an admissible estimator of ω with quadratic loss.

Proof of Lemma 3. In the case $p = 1$ it is easy to see that $Z^1, \ldots, Z^m$ are independent random variables such that Z^i $(i = 1, \ldots, m)$ has the normal distribution $N(\overline{\Theta}^i, \sigma^2)$, where $\sigma^2 \boldsymbol{I} = \operatorname{Var} \boldsymbol{X}$.

To apply the theorem of Stein, we put $\mathcal{X} = R$, $\mathcal{Y} = R^{m-1}$, $X = Z^i$,

$$Y = (Z^1 - \overline{\Theta}^1, \ldots, Z^{i-1} - \overline{\Theta}^{i-1}, Z^{i+1} - \overline{\Theta}^{i+1}, \ldots, Z^m - \overline{\Theta}^m),$$

where $\overline{\Theta}^1, \ldots, \overline{\Theta}^{i-1}, \overline{\Theta}^{i+1}, \ldots, \overline{\Theta}^m$ are arbitrary fixed numbers, $\omega = \overline{\Theta}^i$,

$$p(x, y) = p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{x^2}{2\sigma^2}\Big) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(Z^i)^2}{2\sigma^2}\Big),$$

$$d\nu(y) = \Big(\frac{1}{2\pi\sigma^2}\Big)^{(m-1)/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{\substack{j=1 \\ j \neq i}}^{m} (z^j - \overline{\Theta}^j)^2\Big) \prod_{\substack{j=1 \\ j \neq i}}^{m} dz^j.$$

It is easy to verify that the assumptions of this theorem are satisfied. Thus, it follows from (a) that Z^i is an admissible estimator of $\overline{\Theta}^i$ with quadratic loss.

Since Z^i does not depend on $\overline{\Theta}^j$ for $j \neq i$, Z^i is an admissible estimator of $\overline{\Theta}^i$ for all $\overline{\boldsymbol{\Theta}} \in R^m$.

In the case $p = 2$ it is easy to see that $Z^1, \ldots, Z^m$ are independent random vectors such that Z^i has the two-dimensional normal distribution $N_2(\overline{\Theta}^i, \Sigma)$.

To apply the theorem of Stein in this case we put $\mathcal{X} = R^2$, $\mathcal{Y} = R^{2\times(m-1)}$, $X = Z^i$,

$$Y = (Z^1 - \overline{\Theta}^1, \ldots, Z^{i-1} - \overline{\Theta}^{i-1}, Z^{i+1} - \overline{\Theta}^{i+1}, \ldots, Z^m - \overline{\Theta}^m),$$

where $\bar{\Theta}^1, \ldots, \bar{\Theta}^{i-1}, \bar{\Theta}^{i+1}, \ldots, \bar{\Theta}^m$ are arbitrary fixed elements of R^2, $\omega = \bar{\Theta}^i$,

$$p(x, y) = p(x) = \frac{1}{2\pi} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} x\Sigma^{-1}x^T\right),$$

$$d\nu(y) = \left(\frac{1}{2\pi}\right)^{m-1} |\Sigma|^{-(m-1)/2} \exp\left(-\frac{1}{2} \sum_{\substack{j=1\\ j\neq i}}^{m} (z^j - \bar{\Theta}^j)\Sigma^{-1}(z^j - \bar{\Theta}^j)^T\right) \prod_{\substack{j=1\\ j\neq i}}^{m} dz^j.$$

To verify the condition in (b) it is sufficient to note that $\log^{1+\delta}\|x\|^2$ is integrable with respect to the Lebesgue measure in a neighbourhood of zero, and that all moments of a normal distribution are finite.

Thus, for all $\bar{\Theta} \in R^{2\times m}$, Z^i is an admissible estimator of $\bar{\Theta}^i$ with quadratic loss.

This completes the proof of Lemma 3.

To prove Theorem 2 we use the following proposition which allows us to restrict further considerations to estimators based on the sufficient statistic $\boldsymbol{Z}$:

PROPOSITION. *For every estimator $\lambda(\boldsymbol{X})$ of $\boldsymbol{G\Theta}$, the estimator $\hat{\lambda}(\boldsymbol{Z}) = \mathrm{E}_{\boldsymbol{\Theta}}\big(\lambda(\boldsymbol{X})|\boldsymbol{Z}\big)$ is as good as $\lambda(\boldsymbol{X})$, that is, for all $\boldsymbol{\Theta} \in \Omega$,*

$$V_{\boldsymbol{\Theta}} = \mathrm{E}_{\boldsymbol{\Theta}}\big(\lambda(\boldsymbol{X}) - \boldsymbol{G\Theta}\big)\big(\lambda(\boldsymbol{X}) - \boldsymbol{G\Theta}\big)^T - \mathrm{E}_{\boldsymbol{\Theta}}\big(\hat{\lambda}(\boldsymbol{Z}) - \boldsymbol{G\Theta}\big)\big(\hat{\lambda}(\boldsymbol{Z}) - \boldsymbol{G\Theta}\big)^T$$

is a non-negative definite matrix.

This proposition can be proved in an analogous manner as Rao-Blackwell's theorem.

Proof of Theorem 2. Suppose, for the contrary, that $\lambda(\boldsymbol{Z})$ is a better estimator of $\boldsymbol{G\Theta}$ than $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A}^{-1})\boldsymbol{A}^T\boldsymbol{X}$; thus, for all $\boldsymbol{\Theta} \in \Omega$,

$$B_{\boldsymbol{\Theta}} = \mathrm{E}_{\boldsymbol{\Theta}}\big(\lambda(\boldsymbol{Z}) - \boldsymbol{G\Theta}\big)\big(\lambda(\boldsymbol{Z}) - \boldsymbol{G\Theta}\big)^T - \mathrm{E}_{\boldsymbol{\Theta}}(\tilde{G} - \boldsymbol{G\Theta})(\tilde{G} - \boldsymbol{G\Theta})^T \leqslant 0$$

and $B_{\boldsymbol{\Theta}_0} \neq 0$ for at least one $\boldsymbol{\Theta}_0 \in \Omega$.

It is easy to see that

$$B_{\boldsymbol{\Theta}} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1/2} D_{\boldsymbol{\Theta}} \big(\boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1/2}\big)^T,$$

where

$$D_{\boldsymbol{\Theta}} = \mathrm{E}_{\boldsymbol{\Theta}}\big(\eta(\boldsymbol{Z}) - \bar{\boldsymbol{\Theta}}\big)\big(\eta(\boldsymbol{Z}) - \bar{\boldsymbol{\Theta}}\big)^T - \mathrm{E}_{\boldsymbol{\Theta}}(\boldsymbol{Z} - \bar{\boldsymbol{\Theta}})(\boldsymbol{Z} - \bar{\boldsymbol{\Theta}})^T$$

and

$$\eta(\boldsymbol{Z}) = (\boldsymbol{A}^T\boldsymbol{A})^{1/2}\boldsymbol{G}^{-1}\lambda(\boldsymbol{Z}).$$

Consequently, for all $\boldsymbol{\Theta} \in \Omega$, $D_{\boldsymbol{\Theta}} \leqslant 0$ and $D_{\boldsymbol{\Theta}_0} \neq 0$ for at least one $\boldsymbol{\Theta}_0 \in \Omega$.

Hence, for every $\boldsymbol{\Theta} \in \Omega$, all diagonal elements of $D_{\boldsymbol{\Theta}}$ are non-positive (i.e., $(D_{\boldsymbol{\Theta}})_{ii} \leqslant 0$) and at least one diagonal element, say $D_{\boldsymbol{\Theta}})_{jj}$, is negative for some $\boldsymbol{\Theta}_0 \in \Omega$.

Let Z^j, $\eta^j(\boldsymbol{Z})$ and $\bar{\Theta}^j$ be the j-th rows of $\boldsymbol{Z}$, $\eta(\boldsymbol{Z})$ and $\bar{\boldsymbol{\Theta}}$, respectively. Since, for all $\boldsymbol{\Theta} \in \Omega$, $(D_{\boldsymbol{\Theta}})_{jj} \leqslant 0$ and $(D_{\boldsymbol{\Theta}_0})_{jj} < 0$, for all $\boldsymbol{\Theta} \in \Omega$ there holds

$$\mathrm{E}_{\boldsymbol{\Theta}}\big(\eta^j(\boldsymbol{Z}) - \bar{\Theta}^j\big)\big(\eta^j(\boldsymbol{Z}) - \bar{\Theta}^j\big)^T - \mathrm{E}_{\boldsymbol{\Theta}}(Z^j - \bar{\Theta}^j)(Z^j - \bar{\Theta}^j)^T \leqslant 0$$

with strong inequality for $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$.

Thus $\eta^j(\boldsymbol{Z})$ is a better estimator of $\bar{\Theta}^j$ than Z^j with quadratic loss. This contradicts Lemma 3. Thus the proof of Theorem 2 is complete.

4. Inadmissibility of the least squares estimators in the case $p \geqslant 3$. In this section we show that, in the case $p \geqslant 3$, $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$ can be an inadmissible estimator of $\boldsymbol{G\Theta}$.

Let $\boldsymbol{X}$ be a matrix of form (1), and let assumptions I, II and III be satisfied. Let $F(x_1, \ldots, x_p)$ given in assumption III be the distribution function of a p-dimensional normal distribution such that $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. Let $\boldsymbol{G}$ be a non-degenerate $(m \times m)$-matrix. Then there exist estimators of $\boldsymbol{G\Theta}$ which are better than $\tilde{G} = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{X}$. For example, it can be easily shown that

$$G_1(\boldsymbol{X}) = \boldsymbol{G}(\boldsymbol{A}^T\boldsymbol{A})^{-1/2}\tilde{\boldsymbol{Z}},$$

where $\tilde{\boldsymbol{Z}}$ is a matrix having the same rows as $\boldsymbol{Z} = (\boldsymbol{A}^T\boldsymbol{A})^{-1/2}\boldsymbol{A}^T\boldsymbol{X}$, except the first one Z^1 which is multiplied by $1-(p-2)/Z^1(Z^1)^T$, is better than $\tilde{G}$.

However, if we replace more than one row in $\boldsymbol{Z}$ by Stein's estimator, then we can obtain an estimator which is not as good as $\tilde{G}$.

5. Acknowledgement. I wish to express my thanks to Dr. W. Klonecki for the encouragement during the preparation of this paper.

References

[1] S. M. Fintušal (С. М. Финтушал), *О допустимости оценки Питмэна для многомерного параметра сдвига*, Теория вероятн. прим. 16 (1971), p. 718-723.

[2] A. M. Kagan, *On the estimation theory of location parameter*, Sankhya A 28 (1966), p. 335-3̄72.

[3] — (А. М. Каган), *Теория оценивания для семейств с параметрами сдвига, масштаба и экспонентных*, Труды Матем. Инст. АН СССР 104 (1968), p. 19-88.

[4] — Yu. V. Linnik and C. R. Rao, *On characterization of the normal law based on a property of the sample average*, Sankhya A 27 (1965), p. 405-406.

[5] A. M. Kagan and O. W. Šalaevskiĭ (А. М. Каган и О. В. Шалаевский), *Допустимость оценок наименьших квадратов — исключительное свойство нормального закона*, Математ. заметки 6 (1969), p. 81-89.
[6] Yu. V. Linnik, *Some new investigation on parametric statistics*, European Meeting of Statisticians, 1966.
[7] J. Marcinkiewicz, *Collected papers*, Warszawa 1964, p. 25-29.
[8] C. R. Rao, *Linear statistical inference and its applications*, New York 1966.
[9] A. L. Rukhin (А. Л. Рухин), *Некоторые статистические и вероятностные задачи на группах*, Труды Инст. Стеклова 111 (1970), p. 52-109.
[10] C. Stein, *The admissibility of Pitman estimator of a single location parameter*, Ann. Math. Statist. 30 (1959), p. 970-978.
[11] — and W. James, *Estimation with quadratic loss*, Proc. Fourth Berkeley Sympos. Math. Stat. Prob. 1960, vol. 1, p. 361-379.

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WROCŁAW

Received on 30. 11. 1972

BOGUSŁAWA BEDNAREK-KOZEK (Wrocław)

O ESTYMACJI W WIELOWYMIAROWYM MODELU GAUSSA

STRESZCZENIE

W pracy podano pewną charakteryzację rozkładu normalnego. Rozważając wielowymiarowy model Gaussa-Markowa, udowodniono, że estymatory N.K. (otrzymane metodą najmniejszych kwadratów) pewnych funkcji parametrycznych są dopuszczalne w klasie estymatorów nieobciążonych z macierzową funkcją straty wtedy i tylko wtedy, gdy rozkład próby jest wielowymiarowym rozkładem normalnym.

Ponadto pokazano, że w przypadku p-wymiarowego rozkładu normalnego estymatory N.K. są dopuszczalne z macierzową funkcją straty, gdy $p \leqslant 2$. Gdy $p \geqslant 3$, estymatory N.K. mogą nie być dopuszczalne. W rozdz. 4 pracy podano przykład obciążonego estymatora, lepszego od estymatora N.K. Przykład ten dowodzi, że gdy macierz kowariancji jest znana oraz $p \geqslant 3$, estymatory N.K. są niedopuszczalne.