Sankhyā : The Indian Journal of Statistics
1992, Volume 54, Series A, Pt. 1, pp. 53–73.

CRAMER-RAO TYPE INTEGRAL INEQUALITIES FOR ESTIMATORS OF FUNCTIONS OF MULTIDIMENSIONAL PARAMETER

By B. L. S. PRAKASA RAO
Indian Statistical Institute

SUMMARY. Cramer-Rao type integral inequalities for the integrated risk for estimators of functions of multidimensional parameter are derived extending the work of Borovkov (1984). As an application, a lower bound for the local asymptotic minimax risk of an estimator is obtained when the components of the parameter are orthogonal. Several examples are presented illustrating the results. The problem of estimation of function of mean vector and covariance matrix of a multivariate normal distribution is discussed.

1. Introduction

Let $X_1, X_2, \ldots,$ be independent and identically distributed random variables with values in a measurable space $(\chi, \mathcal{B})$ endowed with a probability measure P_θ, $\theta \in \Theta$ open $\subset R^m$. Suppose that $\{P_\theta, \theta \in \Theta\}$ are dominated by a σ-finite measure μ and $f(\theta, x) = \frac{dP_\theta}{d\mu}(x)$.

We are interested in the problem of estimation of $g(\theta)$, where $g(.)$ is a measurable function defined on R^m, based on the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ when θ is the true but unknown parameter. Let $E_\theta(.)$ denote the expectation under θ. Suppose $g(\theta)$ is a prior probability density for θ such that $S_q \subset \Theta$ where S_q denotes the support of $q(.)$. We denote the expectation over the space $\chi^n \times \Theta$ with respect to the density $f_n(\theta, \boldsymbol{x})\, q(\theta)$ by E. Here

$$f_n(\theta), \boldsymbol{x}) = \pi_{i=1}^{n} f(\theta, x_i). \qquad \ldots \quad (1.1)$$

Our aim is to obtain lower bounds for the integrated risk, namely,

$$R(\theta^*) = E[g(\theta^*) - g(\theta)]^2 \qquad \ldots \quad (1.2)$$

where θ^* is any estimator of θ. For earlier work in this direction, see Borovk (1984) and Borovkov and Sakhanenko (1980) for case $m = 1$ and

Paper received. January 1989; revised May 1990.

AMS (1980) *subject classification.* Primary 62F15, 62H12; secondary 62F12.

Keywords and phrases. Cramer-Rao type integral inequalities; integrated risk; local asymptotic minimax property; multivariate normal distribution.

$g(\theta) = \theta$ and Shemyakin (1987) for the case $m > 1$ and $g(\theta) = \theta$. In fact, Shemyakin (1987) obtained lower bound for the matrix of the integrated risk, namely,

$$R(\theta^*) = E(\theta^* - \theta)(\theta^* - \theta)^t \quad \dots \quad (1.3)$$

where $\boldsymbol{\alpha}^t$ denotes the transpose of column vector $\boldsymbol{\alpha}$.

Section 2 contains Cramer-Rao type integral inequalities for the integrated risk for estimators $g(\theta^*)$ of a parametric function $g(\theta)$ of a multidimensional parameter θ. Some special cases of results in Section 2 are derived in Section 3. As a consequence, a lower bound for the local asymptotic minimax risk for the estimator $g(\theta^*)$ is derived. It was shown that the lower bounds for the integrated risk and the local asymptotic minimax risk are sharp when the components of θ are orthogonal. Several examples illustrating the results derived earlier are presented in Section 5. The problems of estimation of linear function of mean and variance and the estimation of the ratio of mean and variance for a normal distribution and the problem of estimation of ratio of mean of two independent exponential random variables are discussed. The problems of estimation of functions of the form $\boldsymbol{\alpha}^t \beta \boldsymbol{\alpha} + \boldsymbol{\gamma}^t \boldsymbol{V}(\psi)\, \boldsymbol{\delta}$, where $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are known k-dimensional vectors, $\beta\boldsymbol{\alpha}$ and $\boldsymbol{V}(\psi)$ are the mean vector and covariance matrix respectively of a k-variate normal distribution, with β and ψ unknown scalar parameters and $\boldsymbol{V}(.)$ has a known functional form, has been investigated in Section 6. Special case of the problem when $\boldsymbol{V}(\psi) = \boldsymbol{\Delta}\,\psi$ where $\boldsymbol{\Delta}$ is known is studied in detail. Some remarks are made in Section 7.

2. Main result

We first state a well-known result.

Lemma 2.1. *Let* $\boldsymbol{Z} = (Z_1, \dots, Z_m)$ *be a random vector such that* $E(Z_i^2) < \infty$, $1 \leqslant i \leqslant m$ *and* Y *be another random variable such that* $E(Y^2) < \infty$. *Let* $\gamma_i = Cov(Z_i, Y)$, $1 \leqslant i \leqslant m$. *Let* $\boldsymbol{\Sigma}$ *denote the covariance matrix of* $\boldsymbol{Z}$. *Suppose* $\boldsymbol{\Sigma}$ *is positive definite. Then*

$$Var\,(Y) \geqslant \boldsymbol{\gamma}^t\, \boldsymbol{\Sigma}^{-1}\, \boldsymbol{\gamma} \quad \dots \quad (2.1)$$

where $\boldsymbol{\gamma}^t = (\gamma_1, \dots, \gamma_m)$.

In particular, if λ is the largest eigen value of $\boldsymbol{\Sigma}$, then

$$\mathrm{Var}(Y) \geqslant \frac{1}{\lambda} \sum_{i=1}^{m} \gamma_i^2 \geqslant \frac{\sum_{i=1}^{m} \gamma_i^2}{\mathrm{tr}(\boldsymbol{\Sigma})} \quad \dots \quad (2.2)$$

where $tr(\Sigma)$ denotes the trace of the matrix Σ. If Σ is a diagonal matrix with entries σ_i^2, then $\text{Var}(Y) \geqslant \sum_{i=1}^{m} \gamma_i^2/\sigma_i^2$.

Proof. The first part is a consequence of Cauchy-Schwartz inequality and the second part is an easy consequence of properties of positive definite matrices.

We assume that the following regularity conditions hold.

(C1) Let $K_i(\theta, x)$, $1 \leqslant i \leqslant m$ be functions jointly measurable in (θ, x) and absolutely integrable with respect to $\lambda \times \mu^n$ such that

$$\int_{\Theta} K_i(\theta, x)\, d\theta = 0,\ 1 \leqslant i \leqslant m,\ x \in \chi^n. \qquad \ldots \quad (2.3)$$

Hence λ is the Lebesgue measure on R^m. Define

$$C_i(\theta, x) = \frac{K_i(\theta, x)}{f_n(\theta, x)\, q(\theta)},\ 1 \leqslant i \leqslant m,\ x \in \chi^n,\ \theta \in S_q. \qquad \ldots \quad (2.4)$$

(C2) For simplicity, we assume that $S_q = \Theta$. It can be checked that all the following arguments hold if S_q is a proper subset of Θ. Hence $q(\theta) > 0$ for all $\theta \in \Theta = S_q \subset R^m$.

(C3) Let $g(\theta^*)$ be an estimator of $g(\theta)$ and suppose that $g(\theta^*)\, K_i(\theta, x)$ is jointly measurable in (θ, x) and absolutely integrable with respect to $\lambda \times \mu^n$ on $\Theta \times \chi^n$.

(C4) Let $h(\theta)$ be a measurable function of θ such that $f_n(\theta, x)\, h(\theta)$ is differentiable with respect to θ componentwise. Let

$$K_i(\theta, x) = \frac{\partial}{\partial \theta_i}\ [f_n(\theta, x)\, h(\theta)].$$

Suppose $K_i(\theta, x)$ is absolutely integrable with respect to $\lambda \times \mu^n$ on $\Theta \times \chi^n$. Furthermore suppose that, for every $x \in \chi^n$

$$f_n(\theta, x)\, h(\theta) \to 0 \text{ as } |\theta_i| \to \text{boundary of } \Theta_i \text{ for all } i,\ 1 \leqslant i \leqslant m$$

where Θ_i is the range of the i-th component.

(C5) Suppose that differentiation componentwise with respect to θ under the integral sign is valid in the equation

$$\int_{\chi^n} f_n(\theta, x)\, \mu^n(dx) = 1.$$

(C6) $I_1(\theta) = \sum_{i=1}^{m} E_\theta \left[\frac{\partial \log f(X, \theta)}{\partial \theta_i}\right]^2$ is continuous in θ.

Let

$$Y = g(\theta^*) - g(\theta) \text{ and } Z_i = C_i(\theta, X). \qquad \ldots \quad (2.5)$$

Observe that

$$\begin{aligned} E(Z_i) &= E(E_\theta(G_i(\theta, \boldsymbol{X}))) \\ &= E\{\int_{\chi^n} G_i(\theta, x) f_n(\theta, x) \mu^n(dx)]\} \\ &= \int_\Theta \{\int_{\chi^n} G_i(\theta, x) f_n(\theta, x) \mu^n(dx)\} q(\theta)\, d\theta \\ &= \int_\Theta \int_{\chi^n} K_i(\theta, x) \mu^n(dx)\, d\theta \\ &= \int_{\chi^n} [\int_\Theta K_i(\theta, x)\, d\theta]\, \mu^n(dx) = 0 \qquad \ldots (2.6) \end{aligned}$$

by Fubini's theorem. Furthermore

$$E[YZ_i] = E[(g(\theta)^* - g(\theta))\, G_i(\theta, \boldsymbol{X})] = -E[g(\theta)\, G_i(\theta, \boldsymbol{X})] \qquad \ldots (2.7)$$

since, by Fubini's theorem,

$$\begin{aligned} E[g(\theta^*)\, G_i(\theta, \boldsymbol{X})] &= \int_\Theta \int_{\chi^n} g(\theta^*(x))\, G_i(\theta, x) f_n(\theta, x)\, q(\theta)\, \mu^n(dx) d\theta \\ &= \int_\Theta \int_{\chi^n} g(\theta^*(x))\, K_i(\theta, x)\, \mu^n(dx) d\theta \\ &= \int_{\chi^n} g(\theta^*(x)) [\int_\Theta K_i(\theta, x)\, d\theta]\, \mu^n(dx) = 0. \qquad \ldots (2.8) \end{aligned}$$

Note that

$$\text{Cov}(Y, Z_i) = E[YZ_i] - E(Y)E(Z_i) = -E[g(\theta)\, G_i(\theta, \boldsymbol{X})] \qquad \ldots (2.9)$$

from (2.6) and (2.7) and

$$\text{Cov}(Z_i, Z_j) = E(Z_i Z_j) - E(Z_i)\, E(Z_j) = E[G_i(\theta, \boldsymbol{X})\, G_j(\theta, \boldsymbol{X})]. \qquad \ldots (2.10)$$

Hence, by Lemma 2.1, it follows that

$$\text{Var}[Y] \geqslant \boldsymbol{\gamma}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} \qquad \ldots (2.11)$$

where

$$\boldsymbol{\gamma} = \begin{pmatrix} -E[g(\theta)\, G_1(\theta, \boldsymbol{X})] \\ -E[g(\theta)\, G_m(\theta, \boldsymbol{X})] \end{pmatrix}_{m\times 1} \qquad \ldots (2.12)$$

and

$$\boldsymbol{\Sigma} = ((E[G_i(\theta, \boldsymbol{X})\, G_j(\theta, \boldsymbol{X})]))_{m\times m}. \qquad \ldots (2.13)$$

We have now the following result.

Theorem 2.1. *Suppose the conditions (C1) to (C3) hold. Then*

$$E[g(\theta^*)-g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m} (E[g(\theta)\, G_i(\theta, X)])^2}{\sum_{i=1}^{m} E(G_i\, (\theta, X)^2]} + (E[g(\theta^*)-g(\theta)])^2. \qquad \ldots \quad (2.14)$$

Let us now consider a special case of the inequality (2.14). Suppose the conditions (C4) and (C5) hold and (C3) is satisfied for $K_i(\theta, x)$ specified in (C4). Then

$$K_i(\theta, x) = \frac{\partial}{\partial\theta_i} [f_n(\theta, x)\, h(\theta)] \qquad \ldots \quad (2.15)$$

and

$$\int_\theta K_i(\theta, x)\, d\theta = \int_\theta \frac{\partial}{\partial\theta_i} [f_n(\theta, x)\, h(\theta)]\, d\theta$$

$$= \int_{\theta_1\times\ldots\times\theta_{i-1}\times\theta_{i+1}\ldots\times\theta_m} \left\{ \int_{\theta_i} \frac{\partial}{\partial\theta_i} [f_n(\theta, x)\, h(\theta)] d\theta_i \right\} d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_m$$

(by Fubini's theorem)

$$= \int_{\theta_1\times\ldots\times\theta_{i-1}\times\theta_{i+1}\times\ldots\times\theta_m} \left\{ \int_{\theta_i} d_i[f_n(\theta, x) h(\theta)] \right\} d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_m$$

$$\ldots \quad (2.16)$$

where d_i denotes the differential with respect to θ_i (keeping $\theta_1, \ldots, \theta_{i-1}$, $\theta_{i+1} \ldots, \theta_m$ fixed).

Hence, by (C4), it follows that

$$\int_\theta K_i(\theta, x)\, d\theta = 0 \text{ for } 1 \leqslant i \leqslant m \text{ and for every } x \in \chi^*. \qquad \ldots \quad (2.17)$$

Note that

$$G_i\, (\theta, x)$$

$$= \frac{K_i\, (\theta, x)}{f_n(\theta, x)\, q(\theta)} = \frac{\frac{\partial}{\partial\theta_i} [f_n(\theta, x) h(\theta)]}{f_n(\theta, x)\, q(\theta)} = \frac{f_n^{(i)}\, (\theta, x)\, h(\theta) + f_n(\theta, x) h^{(i)}\, (\theta)}{f_n(\theta, x)\, q(\theta)}$$

$$\ldots \quad (2.18)$$

where $f_n^{(i)}\, (\theta, x)$ and $h^{(i)}\, (\theta)$ denote the partial derivatives of $f_n(\theta, x)$ and $h(\theta)$ respectively with respect to θ_i. Hence

$$G_i(\theta, x) = \frac{f_n^{(i)}\, (\theta, x)\, h(\theta)}{f_n(\theta, x)\, q(\theta)} + \frac{h^{(i)}(\theta)}{q(\theta)} = \frac{\partial \log f_n(\theta, x)}{\partial\theta_i} \; \frac{h(\theta)}{q(\theta)} + \frac{h^{(i)}\, (\theta)}{q(\theta)}$$

$$\ldots \quad (2.14)$$

Now

$$E[g(\theta)G_i(\theta, \boldsymbol{X})] = \int_{\Theta} \int_{\chi^n} \{g(\theta)\, G_i(\theta, x)\, f_n(\theta, x)\, q(\theta)\}\ \mu^n(dx)d\theta$$

$$\int_{\Theta} \int_{\chi^n} g(\theta)\, [f_n^{(i)}(\theta, x)\, h(\theta) + f_n(\theta, x)\, h^{(i)}(\theta)]\ \mu^n(dx)d\theta$$

$$= \int_{\Theta} \int_{\chi^n} g(\theta)\ \frac{\partial}{\partial \theta_i} [f_n(\theta, x)\, h(\theta)]\ \mu^n(dx)d\theta$$

$$= -\int_{\Theta} \int_{\chi^n} g^{(i)}(\theta)\, f_n(\theta, x)\, h(\theta)\ \mu^n(dx)d\theta$$

$$= -\int_{\Theta} \int_{\chi^n} g^{(i)}(\theta)\, f_n(\theta, x) h(\theta)\ \mu^n(dx)d\theta$$

$$= -E\left[\frac{g^{(i)}(\theta)h(\theta)}{q(\theta)} \right] \qquad \ldots \quad (2.10)$$

where $g^{(i)}(\theta)$ denotes the derivative of $g(\theta)$ with respect to θ_i. Furthermore

$$E[G_i^2(\theta, \boldsymbol{X})] = E\left[\frac{f_n^{(i)}(\theta, \boldsymbol{X})\, h(\theta)}{f_n(\theta, \boldsymbol{X})\, q(\theta)} \right]^2 + E\left[\frac{h^{(i)}(\theta)}{q(\theta)} \right]^2 + 2E\left[\frac{f_n^{(i)}(\theta, \boldsymbol{X}) h(\theta) h^{(i)}(\theta)}{f_n(\theta, \boldsymbol{X})\, q^2(\theta)} \right]$$

$$= E\left[\frac{f_n^{(i)}(\theta, \boldsymbol{X})}{f_n(\theta, \boldsymbol{X})}\ \frac{h(\theta)}{q(\theta)} \right]^2 + E\left[\frac{h^{(i)}(\theta)}{q(\theta)} \right]^2 \qquad \ldots \quad (2.21)$$

Observe that $E_\theta\left[\frac{f_n^{(i)}(\theta, \boldsymbol{X})}{f_n(\theta, \boldsymbol{X})} \right] = 0$ from (C5). Hence

$$E[G_i^2(\theta, \boldsymbol{X})] = E\left[I_n^{(i)}(\theta)\, \frac{h^2(\theta)}{q^2(\theta)} \right] + E\left[\frac{h^{(i)}(\theta)}{q(\theta)} \right]^2 \text{ where } I_n^{(i)}(\theta) = E_\theta \left[\frac{f_n^{(i)}(\theta, \boldsymbol{X})}{f_n(\theta, \boldsymbol{X})} \right] \qquad \ldots \quad (2.22)$$

Hence we have the following result from Theorem 2.1.

Theorem 2.2. *Suppose the condition (C3) to (C5) hold. Then*

$$E[g(\theta^*) - g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m} \left(E\left[g^{(i)}(\theta)\, h(\theta) q(\theta) \right] \right)^2}{\sum_{i=1}^{m} E\left[I_n^{(i)}(\theta)\, \frac{h^2(\theta)}{q^2(\theta)} \right] + \sum_{i=1}^{m} E\left[\frac{h^{(i)}(\theta)}{q(\theta)} \right]^2} + E(g(\theta^*) - g(\theta)]^2. \qquad \ldots \quad (2.23)$$

In particular,

$$E[g(\theta^*) - g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m} \left(E\left[\frac{g^{(i)}(\theta) h(\theta)}{q(\theta)} \right] \right)^2}{\sum_{i=1}^{m} E\left[I_n^{(i)}(\theta)\, \frac{h^2(\theta)}{q^2(\theta)} \right] + \sum_{i=1}^{m} E\left[\frac{h^{(i)}(\theta)}{q(\theta)} \right]^2} \qquad \ldots \quad (2.24)$$

3. Special cases

Let us now consider further special cases of (2.24) which are useful. In the following discussion, we assume that appropriate special cases of conditions (C2) to (C5) hold.

3.1. Suppose $h(\theta) = \frac{q(\theta)}{I_n(\theta)}$ Note that

$$I_n(\theta) = \sum_{i=1}^{m} I_n^{(i)}(\theta) = \sum_{i=1}^{m} E_\theta \left[\frac{\partial \log f_n(\theta, X)}{\partial \theta_i}\right]^2 = n\, I_1(\theta)$$

under (C5) and hence from (2.24),

$$E[g(\theta^* - g(\theta))]^2 \geqslant \frac{\sum_{i=1}^{m} \left(E\left[\frac{g^{(i)}(\theta)}{I_n(\theta)}\right]\right)^2}{E\left[\frac{1}{I_n(\theta)}\right] + \sum_{i=1}^{m} E\left[\frac{h^{(i)}(\theta)}{q(\theta)}\right]^2}$$

$$= \frac{\frac{1}{n^2} \sum_{i=1}^{m} \left(E\left[\frac{g^{(i)}(\theta)}{I_1(\theta)}\right]\right)^2}{\frac{1}{n} E\left[\frac{1}{I_1(\theta)}\right] + \sum_{i=1}^{m} E\left[\frac{h^{(i)}(\theta)}{q(\theta)}\right]^2} \qquad \ldots \quad (3.1)$$

3.2. Suppose $h(\theta) = q(\theta)$. Let $I_n(\theta) = \sum_{i=1}^{m} I_n^{(i)}(\theta)$. Then

$$E[g(\theta^*) - g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m} (E[g^{(i)}(\theta)])^2}{E[I_n(\theta)] + \sum_{i=1}^{m} E\left[\frac{q^{(i)}(\theta)}{q(\theta)}\right]^2}$$

$$= \frac{\sum_{i=1}^{m} (E[g^{(i)}(\theta)])^2}{n\, E[I_1(\theta)] + \sum_{i=1}^{m} E\left[\frac{\partial \log q(\theta)}{\partial \theta_i}\right]^2} \qquad \ldots \quad (3.2)$$

3.3. Suppose $\epsilon > 0$ exists such that $J = \{\theta : |\theta_i - \theta_{i0}| < \epsilon, 1 \leqslant i \leqslant m\}$ is contained in $\Theta \subset R^{n}$. Then

$$\sup_{\theta \in J} E_\theta[g(\theta^*) - g(\theta)]^2 \geqslant \int_J E_\theta[g(\theta^*) - g(\theta)]^2\, q(\theta) d\theta \qquad \ldots \quad (3.3)$$

where $q(.)$ is a prior on J. Let us choose

$$h(\theta) = q(\theta) = \prod_{i=1}^{m} \left\{\frac{1}{\epsilon} \cos^2 \frac{\pi(\theta_i - \theta_{i0})}{2\epsilon}\right\}, \theta \in J. \qquad \ldots \quad (3.4)$$

We have seen, from results in 3.2, that

$$E[g(\theta^*)-g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m}(E[g^{(i)}(\theta)])^2}{n\,E[I_1(\theta)]+\sum_{i=1}^{m} E\left[\frac{\partial \log q(\theta)}{\partial\theta_i}\right]^2} \qquad \ldots \quad (3.5)$$

It is easy to check that (cf. Borovkov, 1984, p. 189)

$$E\left[\frac{\partial \log q(\theta)}{\partial\theta_i}\right]^2 = \pi^2\,\epsilon^{-2},\ 1 \leqslant i \leqslant m$$

Hence

$$\sup_{\theta\in J} E[g(\theta^*)-g(\theta)]^2 \geqslant \frac{\sum_{i=1}^{m}(E[g^{(i)}(\theta)])^2}{nE[I_1(\theta)]+m\,\pi^2\,\epsilon^{-2}}$$

$$\geqslant \frac{\sum_{i=1}^{m}(E[g^{(i)}(\theta)])^2}{n\sup_{\theta\in J} I_1(\theta)+m\,\pi^2\,\epsilon^{-2}} \qquad \ldots \quad (3.6)$$

In particular,

$$\sup_{\theta\in J} E_\theta[\sqrt{n}\,(g(\theta^*)-g(\theta))]^2 \geqslant \frac{n\sum_{i=1}^{m}(E(g^{(i)}(\theta)])^2}{nE(I_1(\theta))+m\,\pi^2\,\epsilon^{-2}}$$

$$\geqslant \frac{\sum_{i=1}^{m}(E[g^{(i)}(\theta)])^2}{\sup_{\theta\in J} I_1(\theta)+m\pi^2\,\epsilon^{-2}n^{-1}} \qquad \ldots \quad (3.9)$$

Let $\epsilon_n \to 0$ as $n\to\infty$ such that $\epsilon_n^{-2}\,n^{-1}\to 0$ and $J_n = \{\theta : |\theta_i-\theta_{i0}| \leqslant \epsilon_n,\ 1 \leqslant i \leqslant m\}$. For instance $\epsilon_n = n^{-\alpha}$ where $0 < \alpha < 1/2$ will be such a sequence. Note that

$$\sup_{\theta\in J} E_\theta\,[\sqrt{n}(g\,(\theta^*)-g(\theta))]^2 \geqslant \frac{\sum_{i=1}^{m}(E[g^{(i)}(\theta))^2}{\sup_{\theta\in J} I_1(\theta)+m\pi^2\,\epsilon_n{}^{-2}n^{-1}} \qquad \ldots \quad (3.8)$$

and taking limit as $n\to\infty$, we have

$$\liminf_{n\to\infty}\ \sup_{\theta\in J_n} E_\theta\,[\sqrt{n}\,(g(\theta^*)-g(\theta))]^2 \geqslant \frac{\sum_{i=1}^{m}(g^{(i)}(\theta_0)]^2}{I_1(\theta_0)} \qquad \ldots \quad (3.9)$$

The last relation follows from the observation

$$E(g^{(i)}(\theta)] = \int_{J_n} g^{(i)}(\theta)\,q(\theta)d\theta = \frac{1}{\epsilon_n^m}\underbrace{\int\cdots\int}_{J_n} g^{(i)}(\theta)\prod_{i=1}^{m}\cos^2\frac{\pi(\theta_i-\theta_{i0})}{2\epsilon_n}\,d\theta. \quad \ldots \quad (3.10)$$

and, as $n \to \infty$,

$$E(g^{(i)}(\theta)] \to g^{(i)}(\theta_0) \quad \ldots \quad (3.11)$$

since $\epsilon_n \to 0$. Furthermore by (C6),

$$\sup_{\theta \in J_n} I_1(\theta) \to I_1(\theta_0). \quad \ldots \quad (3.12)$$

The lower bound in (3.9) is not sharp as we have taken a weak lower bound in (2.2). However, it is easily computable. The lower bounds obtained in this section as well as throughout this paper are valid for all estimators g^* of $g(\theta)$ and not necessarily for those of the form $g(\theta^*)$.

4. Orthogonal parameters

Suppose the components of $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ are orthogonal in the sense that

$$E_\theta\left[\frac{\partial \log f(\theta, X)}{\partial \theta_i} \quad \frac{\partial \log f(\theta, X)}{\partial \theta_j} \right] = 0, \; 1 \leqslant i \neq j \leqslant m \quad \ldots \quad (4.1)$$

(cf. Cox and Reid, 1987). Let $h(\theta) = q(\theta)$ as in 3.3. Note that, for $i \neq j$,

$$E[G_i(\theta, x)\, G_j(\theta, x)] = E\left[\frac{K_i(\theta, x)\, K_j(\theta, x)}{f_n^2(\theta, x)\, q^2(\theta)} \right]$$

$$= \int_{\chi^n} \int_{\Theta} \frac{\frac{\partial}{\partial \theta_i}[f_n(\theta, x)\, q(\theta)] \frac{\partial}{\partial \theta_j}[f_n(\theta, x)\, q(\theta)]}{f_n(\theta, x)\, q(\theta)} \mu^n(dx) d\theta$$

$$= \int_{\chi^n} \int_{\Theta} \frac{\partial \log [f_n(\theta, x)\, q(\theta)]}{\partial \theta_i} \quad \frac{\partial \log [f_n(\theta, x)\, q(\theta)]}{\partial \theta_j} f_n(\theta, x) q(\theta) \; \mu^n(dx) d\theta = 0$$

from (4.1), (C5), independence of X_i, $1 \leqslant i \leqslant n$ and the choice of $q(\theta)$ as in (3.4).

Hence the matrix Σ defined by (2.13) is a diagonal matrix and it can be checked that

$$E[g(\theta^*) - g(\theta)]^2 \geqslant \sum_{i=1}^{m} \frac{(E[g(\theta)\, G_i(\theta, X)])^2}{E(G_i^2(\theta, X))} + (E[g(\theta^*) - g(\theta)])^2$$

$$\geqslant \sum_{i=1}^{m} \frac{(E[g^{(i)}(\theta)])^2}{E(I_n^{(i)}(\theta)) + E\left[\frac{q^{(i)}(\theta)}{q(\theta)} \right]^2}$$

$$= \sum_{i=1}^{m} \frac{(E[g^{(i)}(\theta)])^2}{nE(I_1^{(i)}(\theta)) + E\left[\frac{\partial \log q(\theta)}{\partial \theta_i} \right]^2} \quad \ldots \quad (4.2)$$

It is easy to see that the above inequality holds for any prior density of θ of the form $q(\theta) = \prod_{i=1}^{m} q_i(\theta_i)$ where $q_i(\theta_i)$ is a prior density of θ_i and $I_1^{(i)}(\theta) = E_\theta \left[\frac{\partial \log f(\theta, X)}{\partial \theta_i}\right]$. The lower bound is sharp here since $\gamma^t \Sigma^{-1} \gamma = \sum_{i=1}^{m} (\gamma_i^2/\sigma_i^2)$ where $\sigma_1^2, \ldots, \sigma_m^2$ denote the diagonal elements of Σ.

Choosing $q(\theta) = \prod_{i=1}^{m} \left\{ \frac{1}{\epsilon} \cos^2 \frac{\pi(\theta_i - \theta_{i0})}{2\epsilon} \right\}$, $\theta \in J$ as before as prior density, it can be checked that

$$\sup_{\theta \in J} E_\theta[g(\theta^*) - g(\theta)]^2 \geqslant \sum_{i=1}^{m} \frac{(E[g^{(i)}(\theta)])^2}{n \sup_{\theta \in J} I_1^{(i)}(\theta) + \pi^2 \epsilon^{-2} m} \qquad \ldots (4.3)$$

Let $\epsilon_n \to 0$ such that $\epsilon_n^{-2} n^{-1} \to 0$ as $n \to \infty$. Then it follows that

$$\liminf_{n \to \infty} \sup_{\theta \in J_n} E\, [\sqrt{n}(g(\theta^*) - g(\theta))]^2 \geqslant \sum_{i=1}^{m} \frac{[g^{(i)}(\theta_0)]^2}{I_1^{(i)}(\theta_0)}, \qquad \ldots (4.4)$$

where $J_n = \{\theta : |\theta_i - \theta_{i0}| \leqslant \epsilon_n, 1 \leqslant i \leqslant m\}$. If the equality is obtained in (4.4) for every $\theta_0 \in \theta$, then $g(\theta^*)$ is a *locally asymptotically minimax* estimator of $g(\theta)$.

5. Examples

We now illustrate our results. Detailed calculations are omitted.

Example 5.1. Let X_i, $1 \leqslant i \leqslant n$ be i.i.d. random variables $N(\mu, \sigma^2)$. It can be checked that μ and σ^2 are orthogonal parameters and a locally asymptotic minimax estimator of $g(\theta) = \mu + b\sigma^2$ where b is a known constant is

$$\delta(X) = \bar{X} + \frac{b}{n+1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

The problem of estimation of linear function of the normal mean and variance has been studied recently by Rukhin (1987) among others. He showed that the estimator $\delta(X)$ is inadmissible for $g(\theta)$ for loss function of the form

$$L(\mu, \sigma, \delta) = (\delta - \mu)^2/\sigma^4.$$

As far as we know, multivariate version of this problem has not been discussed in the literature. We investigate this question in Section 6.

Example 5.2 (*Continuation of Example* 5.1). Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$. Let $\theta = \left(\frac{\mu}{\sigma^2}, \mu^2+\sigma^2\right)$. It is known that this reparametrization also gives an orthogonal parametrization of the normal density (cf. Cox and Reid, 1987). It can be shown that

$$\delta(X) = \frac{\bar{X}}{S^2}$$

where $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is a locally asymptotic minimax estimator of $g(\theta) = \frac{\mu}{\sigma^2}$.

Example 5.3. Suppose X and Y are independent exponential random variables with means λ and ψ respectively. Let $Z = (X, Y)$ and suppose we are interested in the estimation of the ratio ψ/λ based on an i.i.d. sample Z_i, $1 \leqslant i \leqslant n$. A convenient reparametrization of the family of distributions of (X, Y) in terms of orthogonal parameters θ_1, θ_2 is given by $\lambda = \theta_1 \theta_2^{-\frac{1}{2}}$ and $\psi = \theta_1 \theta_2^{\frac{1}{2}}$. Then

$$g(\theta_1, \theta_2) = \theta_2 = \psi/\lambda$$

and it can be proved that

$$\delta(\boldsymbol{X}, \boldsymbol{Y}) = \left(\frac{n-1}{n}\right) \left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \right)$$

is a locally asymptotic minimax estimator of ψ/λ.

6. Estimation problem for multivariate normal distribution

Let us now consider a multidimensional version of Example 5.1. As far as we are aware, this problem has not been discussed in the literature.

Suppose $\boldsymbol{X}_i$, $1 \leqslant i \leqslant n$ are i.i.d. k-dimensional random vectors with multivariate normal distribution with mean vector $\beta\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{V}(\psi)$ where β and ψ are unknown scalar parameters and $\boldsymbol{\alpha}$ and $\boldsymbol{V}(.)$ are known k-dimensional vector and $k \times k$ matrix respectively. Further suppose that β and ψ are not functionally related. Then it follows that β and ψ are orthogonal parameters (cf. Cox and Reid, 1987). We are interested in the problem of estimating

$$g(\beta, \psi) = \beta\, \boldsymbol{\alpha}^t \boldsymbol{\alpha} + \boldsymbol{\gamma}^t \boldsymbol{V}(\psi) \boldsymbol{\delta} \qquad \ldots \quad (6.1)$$

where $\boldsymbol{\alpha}, \boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are known k-dimensional vectors. Note that this is a generalization of the problem of estimating linear function of mean μ and variance σ^2 in Example 5.1 to the multivariate case. Let $\theta = (\beta, \psi)$. Observe that

$$g^{(1)}(\theta) = \boldsymbol{\alpha}^t \boldsymbol{a} \text{ and } g^{(2)}(\theta) = \boldsymbol{\gamma}^t \boldsymbol{V}'(\psi)\boldsymbol{\delta} \quad \ldots \quad (6.2)$$

where $\boldsymbol{V}'(\psi)$ denotes the matrix obtained by taking the derivatives of elements of $\boldsymbol{V}(\psi)$ with respect to ψ and forming the matrix of such derivatives. Here we assume that the elements of $\boldsymbol{V}(\psi)$ are differentiable with respect to ψ.

Observing that β and ψ are orthogonal parameters, from results proved earlier, it follows that

$$E[g(\theta^*)-g(\theta)]^2 \geqslant \sum_{i=1}^{2} \frac{(E[g^{(i)}(\theta)])^2}{n\, E(I_1^{(i)}(\theta))+E\left[\dfrac{\partial \log q(\theta)}{\partial\theta_i}\right]^2} \quad \ldots \quad (6.3)$$

where $\theta_1 = \beta$ and $\theta_2 = \psi$,

$$I_1^{(i)}(\theta) = E_\theta\left[\frac{\partial \log f(\boldsymbol{X}, \theta)}{\partial \theta_i}\right]^2, i = 1, 2, \quad \ldots \quad (6.4)$$

$q(\theta)$ is a prior density of θ as given by (3.4) with $m = 2$ and

$$f(\boldsymbol{x}, \theta) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\, \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad \ldots \quad (6.5)$$

with $\boldsymbol{\mu} = \beta\, \boldsymbol{a}$ and $\boldsymbol{\Sigma} = \boldsymbol{V}(\psi)$. Furthermore,

$$\liminf_{n\to\infty} \sup_{\theta\epsilon J_n} E_\theta[\sqrt{n}(g(\theta^*)-g(\theta))]^2 \geqslant \sum_{i=1}^{2} \frac{[g^{(i)}(\theta_0)]^2}{I_1^{(i)}(\theta_0)}$$

$$= \frac{(\boldsymbol{\alpha}^t\boldsymbol{a})^2}{I_1^{(1)}(\theta_0)} + \frac{[\boldsymbol{\gamma}^t\boldsymbol{V}'(\psi)\boldsymbol{\delta}]^2}{I_1^{(2)}(\theta_0)} \quad \ldots \quad (6.6)$$

where $J_n = \{\theta : |\theta_i-\theta_{i0}| \leqslant \epsilon_n, i = 1, 2\}$ and $\epsilon_n^{-2}\, n^{-1} \to 0$ as $n\to\infty$. In order to obtain the lower bound in a more explicit form, we have to compute $I_1^{(i)}(\theta), i = 1, 2$. Note that

$$\log f(\boldsymbol{x}, \theta) = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{x}-\boldsymbol{\mu})^t\, \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

$$= -\frac{k}{2} \log(2\pi) + \frac{1}{2} \log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2} (\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \quad \ldots \quad (6.7)$$

It can be checked that

$$\frac{\partial \log f}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \quad \dots \quad (6.8)$$

and hence

$$\frac{\partial \log f}{\partial \beta} = \boldsymbol{a}^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \quad \dots \quad (6.9)$$

Therefore

$$I_1^{(1)}(\theta) = E_\theta \left[\frac{\partial \log f}{\partial \beta}\right]^2 = E_\theta[\boldsymbol{a}^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}\boldsymbol{a}]$$

$$= \boldsymbol{a}^t \boldsymbol{\Sigma}^{-1} E_\theta[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^t] \boldsymbol{\Sigma}^{-1} \boldsymbol{a}$$

$$= \boldsymbol{a}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \boldsymbol{a} = \boldsymbol{a}^t \boldsymbol{\Sigma}^{-1}\boldsymbol{a} = \boldsymbol{a}^t V(\psi)^{-1} \boldsymbol{a}. \quad \dots \quad (6.10)$$

Computation of $I_1^{(2)}(\theta)$ is much more difficult. It is easy to check that (cf. Anderson, 1958, p. 46)

$$\frac{\partial \text{ lof } f}{\partial \boldsymbol{\Sigma}^{-1}} = \frac{1}{2}(\boldsymbol{\Sigma}-(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^t) = \frac{1}{2}(\boldsymbol{\Sigma}-\boldsymbol{Z}) \quad \dots \quad (6.11)$$

where $\boldsymbol{Z} = (\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^t$.

Since $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = I$, it follows that

$$\boldsymbol{\Sigma}\frac{d\boldsymbol{\Sigma}^{-1}}{d\psi} + \frac{d\boldsymbol{\Sigma}}{d\psi}\boldsymbol{\Sigma}^{-1} = 0$$

and hence

$$[d\,\boldsymbol{\Sigma}^{-1}/d\,\psi] = -\boldsymbol{\Sigma}^{-1}[d\,\boldsymbol{\Sigma}/d\psi]\,\boldsymbol{\Sigma}^{-1}. \quad \dots \quad (6.12)$$

Therefore

$$\frac{\partial \log f}{\partial \psi} = \left[\text{Vec}\left(\frac{\partial \log f}{\partial \boldsymbol{\Sigma}^{-1}}\right)\right]^t \text{Vec}\left[\frac{d\boldsymbol{\Sigma}^{-1}}{d\psi}\right]$$

$$= -\frac{1}{2}[\text{Vec}(\boldsymbol{\Sigma}-\boldsymbol{Z})]^t \text{Vec}\left(\boldsymbol{\Sigma}^{-1}\frac{d\boldsymbol{\Sigma}}{d\psi}\boldsymbol{\Sigma}^{-1}\right)$$

$$= \frac{1}{2}\text{ tr}\left[(\boldsymbol{Z}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\frac{d\boldsymbol{\Sigma}}{d\psi}\boldsymbol{\Sigma}^{-1}\right] \quad \dots \quad (6.13)$$

by the relation $(\text{vec } \boldsymbol{A})^t \text{ vec } \boldsymbol{B} = \text{tr}(\boldsymbol{AB})$. Hence

$$\frac{\partial \log f}{d\psi} = \frac{1}{2}\text{ tr}[(\boldsymbol{Z}-\boldsymbol{V})\boldsymbol{V}^{-1}\boldsymbol{V}'\boldsymbol{V}^{-1}] \quad \dots \quad (6.14)$$

where $\boldsymbol{V}$ stands for $\boldsymbol{\Sigma} = \boldsymbol{V}(\psi)$ and $\boldsymbol{V}' = \frac{d\boldsymbol{\Sigma}}{d\psi} = \boldsymbol{V}'(\psi)$. Since $E_\theta(\boldsymbol{Z}) = \boldsymbol{V}$, it follows that

$$E_\theta\left[\frac{\partial \log f}{\partial \psi}\right] = 0 \qquad \dots \quad (6.15)$$

$$I_1^{(2)}(\theta) = E_\theta\left[\frac{\partial \log f}{\partial \psi}\right]^2 = \mathrm{Var}_\theta\left[\frac{\partial \log f}{\partial \psi}\right] \qquad \dots \quad (6.16)$$

$$= \mathrm{Var}_\theta\left\{\frac{1}{2}\,\mathrm{tr}\,[(\boldsymbol{Z}-\boldsymbol{V})\boldsymbol{V}^{-1}\boldsymbol{V}'\boldsymbol{V}^{-1}]\right\} = \frac{1}{4}\,\mathrm{Var}_\theta\,[\mathrm{tr}\,(\boldsymbol{Z}\boldsymbol{V}^{-1}\,\boldsymbol{V}'\,\boldsymbol{V}^{-1})]$$

$$= \frac{1}{4}\,\mathrm{Var}_\theta\,\{\mathrm{tr}\,[\boldsymbol{V}^{\frac{1}{2}}\,\boldsymbol{V}^{-\frac{1}{2}}\,\boldsymbol{Z}\,\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{V}'\,\boldsymbol{V}^{-1}\}$$

$$= \frac{1}{4}\,\mathrm{Var}_\theta\,\{\mathrm{tr}\,\boldsymbol{V}^{\frac{1}{2}}\,\boldsymbol{W}\,\boldsymbol{V}^{\frac{1}{2}}\,\boldsymbol{V}^{-1}\,\boldsymbol{V}'\boldsymbol{V}^{-1}\}] = \frac{1}{4}\,\mathrm{Var}_\theta\,[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{W})]$$

where $\boldsymbol{A} = \boldsymbol{V}^{-\frac{1}{2}}\,\boldsymbol{V}'\boldsymbol{V}^{-\frac{1}{2}}$ and $\boldsymbol{W}$ has standard Wishart distribution. Note that

$$\boldsymbol{A}\boldsymbol{W} = \boldsymbol{H}^t\,\boldsymbol{H}\,\boldsymbol{A}\,\boldsymbol{H}^t\,\boldsymbol{H}\,\boldsymbol{W}$$

where $\boldsymbol{H}$ is an orthogonal matrix such that $\boldsymbol{H}\;\boldsymbol{A}\,\boldsymbol{H}^t$ is a diagonal matrix Λ. Hence

$$\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{W}) = \mathrm{tr}\,(\boldsymbol{H}^t\,\boldsymbol{H}\,\boldsymbol{A}\,\boldsymbol{H}^t\,\boldsymbol{H}\,\boldsymbol{W}) = \mathrm{tr}(\boldsymbol{H}\,\boldsymbol{A}\,\boldsymbol{H}^t\,\boldsymbol{H}\,\boldsymbol{W}\,\boldsymbol{H}^t)$$

$$= \mathrm{tr}\,(\Lambda\boldsymbol{H}\,\boldsymbol{W}\,\boldsymbol{H}^t). \qquad \dots \quad (6.17)$$

Since $\boldsymbol{H}$ is an orthogonal matrix and $\boldsymbol{W}$ has the standard Wishart distribution $\boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^t$ has also the standard Wishart distribution. Therefore

$$\mathrm{Var}_\theta\,[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{W})] = \mathrm{Var}_\theta\,[\mathrm{tr}\,\Lambda\,\widetilde{\boldsymbol{W}})] \qquad \dots \quad (6.18)$$

where $\widetilde{\boldsymbol{W}}$ has the standard Wishart distribution. Hence

$$\mathrm{Var}_\theta\,[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{W})] = \mathrm{Var}_\theta\left[\sum_{i=1}^{k}\lambda_i\,\widetilde{W}_{ii}\right] \qquad \dots \quad (6.19)$$

where $\lambda_1, \dots, \lambda_n$ are the diagonal elements in Λ which are the eigen values of $\boldsymbol{A}$ and W_{ii}, $1 \leqslant i \leqslant k$ are independent random variables each with Chi-square distribution with 1 degree of freedom. This in turn proves that

$$\mathrm{Var}_\theta\,[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{W})] = 2\sum_{i=1}^{k}\lambda_i^2$$

$$= 2\;\mathrm{tr}\,(\boldsymbol{A}^2)$$

$$= 2\;\mathrm{tr}\,(\boldsymbol{V}^{-\frac{1}{2}}\,\boldsymbol{V}'\boldsymbol{V}^{-\frac{1}{2}}\,\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{V}'\boldsymbol{V}^{-\frac{1}{2}})$$

$$= \mathrm{tr}\,[(\boldsymbol{V}^{-1}\boldsymbol{V}')^2]. \qquad \dots \quad (6.20)$$

Relations (6.16) and (6.20) show that

$$I_1^{(2)}(\theta) = E_\theta\left[\frac{\partial \log f}{\partial \psi}\right]^2 = \frac{1}{2}\,\mathrm{tr}[(V^{-1}V')^2] \qquad \dots \quad (6.21)$$

Combining (6.2), (6. 6), (6.10) and (6.21). we have

$$\liminf_{n\to\infty} \sup_{\theta \in J_n} E_\theta[\sqrt{n}(g(\theta^*)-g(\theta))]^2$$

$$\geqslant \frac{(\alpha^t a)^2}{a^t V(\psi_0)^{-1} a} + \frac{2[\gamma^t V'(\psi_0)\,\delta]^2}{\mathrm{tr}[(V(\psi_0)^{-1}V'(\psi_0))]^2} \qquad \dots \quad (6.22)$$

where $J_n = \{\theta : |\theta_i - \theta_{i0}| \leqslant \epsilon_n, i = 1, 2\}$ and $\epsilon_n^{-2}\, n^{-1} \to 0$ as $n \to \infty$.

As a special case, suppose $V(\psi)$ is linear in ψ in the sense that $V(\psi) = \Delta\psi + \Gamma$, where Γ is a known non-negative definite matrix, Δ is a known positive definite matrix and $\psi > 0$. Then $V'(\psi) = \Delta$ and (6.22) reduces to

$$\liminf_{n\to\infty} \sup_{\theta \in J_n} E_\theta[\sqrt{n}(g(\theta^*)-g(\theta))]^2$$

$$\geqslant \frac{(\alpha^t a)^2}{a^t(\Delta\psi_0+\Gamma)^{-1} a} + \frac{2(\gamma^t \Delta\,\delta)^2}{\mathrm{tr}((\Delta\psi_0+\Gamma)^{-1}\Delta)^2} \qquad \dots \quad (6.23)$$

If Δ is positive definite and $\Gamma = 0$, then $V(\psi) = \Delta\psi$ and (6.22) reduces to

$$\liminf_{n\to\infty} \sup_{\theta \in J_n} E_\theta[\sqrt{n}(g(\theta^*)-g(\theta))]^2$$

$$\geqslant \psi_0 \frac{(\alpha^t a)^2}{a^t \Delta^{-1} a} + \frac{2(\gamma^t\Delta\,\delta)^2\,\psi_0^2}{\mathrm{tr}\,(\Delta^{-1}\Delta)^2}$$

$$= \psi_0 \frac{(\alpha^t a)^2}{a^t \Delta^{-1} a} + \frac{2(\gamma^t\Delta\delta)^2\,\psi_0^2}{k} \qquad \dots \quad (6.24)$$

A natural estimator of

$$g(\theta) = g(\beta, \psi) = \beta\,\alpha^t a + \psi\,\gamma^t \Delta \qquad \dots \quad (6.25)$$

is

$$g_1(\theta^*) = a^t\,\bar{X} + C\,\gamma^t\,\tilde{A}\,\delta \qquad \dots \quad (6.26)$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \dots \quad (6.27)$$

$$\tilde{A} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t, \qquad \dots \quad (6.28)$$

and C is a constant chosen so as to minimize the mean square error of $g_1(\theta^*)$. Note that $\bar{X}$ and $\frac{n\tilde{A}}{n-1}$ are unbiased estimators of μ and Σ. Furthermore $\bar{X}$ and $\tilde{A}$ are the maximum likelihood estimators of μ and Σ for $n > k$. Since X is an inadmissible estimator of μ when $k \geqslant 3$, it might be better to use the James-Stein type estimator of the form

$$\hat{\mu} = (I_k - \frac{r(F_n)}{F_n} \Delta^{-1}) \bar{X} \quad \ldots \quad (6.29)$$

as an estimator of μ where

(a) $$r(z) = \min(k-2, z), 0 \leqslant z < \infty,$$

(b) $$F_n = \frac{m \bar{X}^t \Delta^{-2} \bar{X}}{s_n^2},$$

and

(c) $$s_n^2 = [(n-1)k+2]^{-1} \sum_{j=1}^{n} (X_j-\bar{X})^t \Delta^{-1}(X_j-\bar{X}) \quad \ldots \quad (6.30)$$

as suggested by Berger (1976) or a further improvement given in Nickerson (1988). In such an event

$$g_2(\theta^*) = \alpha^t\hat{\mu}+d\, \gamma^t \tilde{A}\, \delta \quad \ldots \quad (6.31)$$

may be chosen as an estimator of $g(\theta)$ where d is a suitable constant. It is difficult to compute the mean square error of $g_2(\theta^*)$ in view of the fact that $\hat{\mu}$ and $\tilde{A}$ are dependent random vectors. However, one can compute the mean square error of $g_1(\theta^*)$ given by (6.26) using the fact that $\bar{X}$ and $\tilde{A}$ are independent random vectors. In fact

$$E_\theta(g_1(\theta^*)] = \alpha^t E_\theta(\bar{X})+C\gamma^t E_\theta(\tilde{A})\delta = \alpha^t\beta\, \alpha+C\psi\gamma^t\Delta\, \delta \left(\frac{n-1}{n}\right) \quad \ldots \quad (6.32)$$

and

$$E_\theta[g_1(\theta^*)-g(\theta)]^2 = \mathrm{Var}_\theta[g_1(\theta^*)]+(E_\theta(g_1(\theta^*)-g(\theta)))^2$$

$$= \mathrm{Var}_\theta[g_1(\theta^*)]+(\gamma^t\Delta\, \delta)^2 \left(\frac{C(n-1)}{n} -1\right)^2\psi^2. \quad \ldots \quad (6.33)$$

But

$$\mathrm{Var}_\theta(g_1(\theta^*)) = \alpha^t\, \mathrm{Covar}_\theta(\bar{X})\alpha+C^2\, \mathrm{Var}_\theta(\gamma^t\tilde{A}\delta) \text{ (by independent of } \bar{X} \text{ and } \tilde{A}]$$

$$= \frac{\psi}{n}\, \alpha^t\Delta\, \alpha+C^2\, \mathrm{Var}_\theta(\gamma^t\tilde{A}\, \delta). \quad \ldots \quad (6.34)$$

Let

$$S = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t. \quad \ldots \quad (6.35)$$

Note that

$$\text{Var}_\theta (\gamma^t \tilde{A} \delta) = \frac{1}{n^2} \text{Var}_\theta (\gamma^t S \delta)$$

$$= \frac{1}{n^2} \text{Var}_\theta [\text{tr}(\gamma^t S \delta)]$$

$$= \frac{1}{n^2} \text{Var}_\theta (\text{tr}(\delta \gamma^t S)]$$

$$= \frac{1}{n^2} \text{Var}_\theta [\text{tr}(BS)] \quad \ldots \quad (6.36)$$

where $B = \delta \gamma^t$. Note that B is not a symmetric matrix. It can be checked that

$$\text{Var}_\theta (\text{tr}(DS)) = 2(n-1) \text{tr}(D \Sigma D \Sigma) \quad \ldots \quad (6.37)$$

for any symmetric matrix D and

$$\text{Var}_\theta (\text{tr}(BS)] = \text{Var}_\theta \left[\text{tr}\left\{\left(\frac{B+B^t}{2}\right) S\right\} \right]$$

$$= 2(n-1) \text{tr}\left[\left(\frac{B+B^t}{2}\right)\Sigma \left(\frac{B+B^t}{2}\right)\Sigma \right]$$

$$= 2(n-1) \tfrac{1}{2}[\text{tr}(B \Sigma B \Sigma)+\text{tr}(B \Sigma B^t \Sigma)]. \quad \ldots \quad (6.38)$$

Combining (6.36) to (6.38) and observing that $\Sigma = \Delta\psi$, we have

$$\text{Var}_\theta (\gamma^t \tilde{A} \delta) = \frac{(n-1)}{n^2} \psi^2 [\text{tr}(\delta\gamma^t \Delta \delta\gamma^t \Delta)+\text{tr}(\delta\gamma^t \Delta \gamma\delta^t \Delta)]. \quad \ldots \quad (6.39)$$

Hence

$$\text{Var}_\theta (\gamma^t \tilde{A} \delta) = \frac{(n-1)}{n^2} \psi^2 [\text{tr} (\gamma^t \Delta\delta \gamma^t \Delta\delta)+\text{tr}(\gamma^t \Delta \gamma \delta^t \Delta\delta)]$$

$$= \frac{(n-1)}{n^2} \psi^2 [(\gamma^t \Delta\delta)^2+(\gamma^t \Delta\gamma) (\delta^t \Delta\delta)]. \quad \ldots \quad (6.40)$$

Therefore, from (6.34), it follows that

$$\text{Var}_\theta [g_1(\theta^*)] = \frac{\psi}{n} (\alpha^t \Delta \alpha)+ \frac{C^2(n-1)}{n^2}\psi^2 [(\gamma^t \Delta\delta)^2+(\gamma^t \Delta\gamma) (\delta^t \Delta\delta)]. \quad \ldots \quad (6.41)$$

Combining (6.33) to (6.41), we have

$$E_\theta\,[g_1(\theta^*)-g(\theta)]^2 = \frac{\psi}{n}\,(\alpha^t\,\Delta\alpha)+\frac{C^2(n-1)}{n^2}\,\psi^2[(\gamma^t\,\Delta\delta)^2+(\gamma^t\,\Delta\gamma)\,(\delta^t\,\Delta\delta)]$$

$$+\psi^2\,(\gamma^t\,\Delta\delta)^2\left(\frac{C(n-1)}{n}-1\right)^2$$

$$=\frac{\psi}{n}\,(\alpha^t\,\Delta\alpha)+\psi^2(\gamma\,\Delta\delta)^2\left\{\frac{C^2(n-1)^2}{n^2}-\frac{2C(n-1)}{n}+1\right\}$$

$$+\psi^2\,\frac{C^2\,(n-1)}{n^2}\,[(\gamma^t\,\Delta\delta)^2+(\gamma^t\,\Delta\gamma)\,(\delta^t\,\Delta\delta)] \qquad \ldots \quad (6.42)$$

The last expression as a function of C is minimized when $C = C^*$ given by

$$(\gamma^t\Delta\delta)^2\left\{\frac{2C^*(n-1)^2}{n^2}-\frac{2(n-1)}{n}\right\} = -\frac{2C^*(n-1)}{n^2}\,[(\gamma^t\Delta\delta)^2+(\gamma^t\Delta\gamma)\,(\delta^t\Delta\delta)].$$

$$\ldots \quad (6.43)$$

This implies

$$\frac{C^*(n-1)}{n^2}\,\{(n-1)\,(\gamma^t\Delta\delta)^2+(\gamma^t\Delta\delta)^2+(\gamma^t\Delta\gamma)\,(\delta^t\Delta\delta)\} = \left(\frac{n-1}{n}\right)(\gamma^t\Delta\delta)^2 \quad \ldots \quad (6.44)$$

or equivalently

$$C^* = \frac{n(\gamma^t\Delta\delta)^2}{n(\gamma^t\Delta\delta)^2+(\gamma^t\Delta\gamma)\,(\delta^t\Delta\delta)} \qquad \ldots \quad (6.45)$$

Let

$$g_3(\theta^*) = \alpha^t\overline{X}+C^*(\gamma^t\,\widetilde{A}\,\delta) \qquad \ldots \quad (6.46)$$

Then $g_3(\theta^*)$ is an estimator of the type (6.26) of $g(\theta)$ minimizing the mean square error and the mean square error of $g_3(\theta^*)$ is given by

$$E_\theta[g_3(\theta^*)-g(\theta)]^2 = \frac{\psi}{n}\,(\alpha^t\,\Delta\,\alpha)+\psi^2(\gamma^t\,\Delta\,\delta)^2$$

$$+\psi^2(\gamma^t\,\Delta\delta)^2\left\{\frac{C^{*2}(n-1)^2}{n^2}-\frac{2C^*(n-1)}{n}\right\}$$

$$+\psi^2C^{*2}\,\frac{2(n-1)}{n^2}\,\{(\gamma^t\,\Delta\,\delta)^2+(\gamma^t\,\Delta\,\gamma)\,(\delta^t\,\Delta\,\delta)\}$$

$$=\frac{\psi}{n}\,(\alpha^t\,\Delta\,\alpha)+\psi^2(\gamma^t\,\Delta\,\delta)^2-C^*\,\frac{(n-1)}{n}\,\psi^2(\gamma^t\,\Delta\,\delta)^2$$

$$=\frac{\psi}{n}(\alpha^t\,\Delta\,\alpha)+\psi^2\,(\gamma^t\,\Delta\,\delta)^2\left\{1-\frac{C^*(n-1)}{n}\right\} \qquad \ldots \quad (6.47)$$

where C^* is as defined by (6.45). Hence

$$\lim_{n \to \infty} E_\theta[\sqrt{n}(g_3(\theta^*)-g_3(\theta))]^2$$

$$= \psi(\alpha^t \Delta \alpha)+\psi^2\{(\gamma^t \Delta \gamma)(\delta^t \Delta \delta)+(\gamma^t \Delta \delta)^2\}. \quad \ldots \quad (6.48)$$

Obviously this limit is not the same as the lower bound in (6.24) when $\psi = \psi_0$. Hence the estimator $g_3(\theta\ )$ of $g(\theta)$ is not locally asymptotic minimax eventhough it is optimum in the sense of minimizing the mean square error. This is unlike the result in Example 5.1 where the one-dimensional case was discussed.

However, in view of the regression structure in the special case (X is $N_k(\beta\, \alpha, \Delta\psi)$ with known α and Δ), one can obtain an estimator which attains the lower bound (6.24) as shown by Dr. C. G. Bhattacharya in a private communication. In fact

$$g_4(\theta^*) = \hat{\beta}\, \alpha^t \alpha+(\gamma^t \Delta \delta)\hat{\psi}$$

where $\hat{\beta}$ is the regression estimator of β and $\hat{\psi}$ is the estimator of ψ based on the error sum of squares is a locally asymptotic minimax estimator of

$$g(\theta) = \beta\, \theta^t \alpha+(\gamma^t \Delta \delta)\, \psi.$$

7. Remarks

After the original version of this paper was submitted for publication, Professor J. K. Ghosh has informed the author that he and Dr. S. N. Joshi have extended the Borovkov's inequality to the special case $g(\theta) = \theta_i$ in an unpublished note in 1983. Babrovsky *et al.* (1987) discussed global Cramer-Rao type bounds for the estimation error of a parameter in a Bayesian set-up both in the one-dimensional as well as in multidimensional case. In the one-dimensional case, they obtain the inequality

$$E(\theta-E(\theta\,|\,X))^2 \geqslant \frac{(Eh(\theta))^2}{E\left[\frac{d}{d\theta}[h(\theta)L_n(X,\ \theta)]/L_n(X,\theta)\right]^2} \quad \ldots \quad (7.1)$$

for any integrable $h(.)$ (see Equation (24) of Babrovsky *et al.* (1987)). The corresponding inequality of Borovkov (1984) is

$$E(\theta-E(\theta\,|\,X))^2 \geqslant \frac{\left(E\frac{h(\theta)}{q(\theta)}\right)^2}{E\left[\frac{d}{d\theta}(L_n(X,\theta)\, h(\theta))/L_n(X,\theta)\, q(\theta)\right]^2} \quad \ldots \quad (7.2)$$

where $q(\theta)$ is the prior density for θ and $h(\theta)$ is a suitable function of θ. Here $L_n(X, \theta)$ is the joint density of $X = (X_1, \ldots, X_n)$ given θ. If $h(\theta) \equiv 1$ and X_i, $1 \leqslant i \leqslant n$ are i.i.d., then the inequality (7.1) reduces to

$$E(\theta - E(\theta \mid X))^2 \geqslant \frac{1}{nE(I(\theta))} \quad \ldots \quad (7.3)$$

where $I(\theta)$ is the Fisher information. Choosing $h(\theta) = \frac{q(\theta)}{I(\theta)}$, the inequality (7.2) reduces to

$$E(\theta - E(\theta \mid X))^2 \geqslant \frac{J^2}{nJ+H} \geqslant \frac{J}{n} - \frac{H}{n^2} \quad \ldots \quad (7.4)$$

for some $H > 0$. From the elementary inequality

$$[E(X)]^{-1} \leqslant E(X^{-1})$$

for a positive random variable X, it follows that the inequality (7.4) gives a better lower bound than that given by (7.3) upto terms of order $o(n^{-1})$. One might get a sharper lower bound than the one given by (7.3) by choosing a suitable function h in (7.1) as pointed out by Babrovsky *et al.* (1987, p. 1428). Similar comments are in order for the multidimensional case between Propositions 2 and 4 of Babrovsky *et al.* (1987) and Lemma 2 of Shemyakin (1987).

The lower bound given in (4.4) is obtained as an application of the Cramer-Rao type integral inequality for the special case when the components of the parameter are orthogonal, the observations are independent and identically distributed (i.i.d.) and the loss function is the squared error loss. However stronger results can be derived for general parameters and for subconvex loss functions when the observations need not be i.i.d. following the general theory developed by Millar (1981, Chapter VII, Theorem 2.6). This was shown by Samanta (1990) in a private communication to the author.

Apart from their applications to problems in checking local asymptotic minimaxity, integral inequalities are interesting in their own way (cf. Babrovsky *et al.* (1987)). It would be interesting to extend these integral inequalities to the non-regular case. An attempt in this direction was made by Babrovsky *et al.* (1987). However their discussion is incomplete.

Acknowledgement. It is a pleasure to thank Dr. C. G. Bhattacharya for his help in computations in Section 6. Thanks are due to Professor J. K. Ghosh for his pertinent comments on the earlier versions and for bringing the papers by Millar (1981), Babrovsky *et al.* (1987) and his own unpublished work with Dr. S. N. Joshi to the author's attention.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Analysis*, Wiley, New York.

Babrovsky, B. Z., Mayer-Wolf, E. and Zakai, M. (1987). Some classes of global Cramer-Rao bounds. *Ann. Statist.*, **15**, 1421-1438.

Berger, J. (1976). Admissibility, minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.*, **4**, 223-226.

Borovkov, A. A. (1984). *Mathematical Statistics*, Nauka, Moscow (In Russian).

Borovkov, A. A. and Sakhanenko, A. I. (1980). On estimates for the average quadratic risk. *Probab. Math. Statist.*, **1**, 185-195, (In Russian).

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. B*, **49**, 1-39.

Millar, P. W. (1981). The minimax principle in asymptotic statistical theory, In *Lecture Notes in Mathematics*, **976**, 75-265 (Ed. P. L. Hennequin) Springer-Verlag, Berlin.

Nickerson, D. (1988). Dominance of the positive part version of the James-Stein estimator. *Statistics and Probability Letters*, **7**, 97-103.

Rukhin, A. L. (1987). Estimating a linear function of the normal mean and variance. *Sankhyā*, Ser. A, **49**, 72-77.

Samanta, T. (1990). Private communication.

Shemyakin, A. E. (1987). Rao-Cramer type integral inequalities for estimates of a vector parameter. *Theory Probab. Appl.*, **32**, 426-434.

Indian Statistical Institute
7 S.J.S. Sansanwal Marg
New Delhi-110016
India.