Journal of Multivariate Analysis 100 (2009) 2296–2304

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Generalized Bayes minimax estimation of the normal mean matrix with unknown covariance matrix

Hisayuki Tsukuma

Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan

ARTICLE INFO

Article history:
Received 7 May 2008
Available online 3 May 2009

AMS 2000 subject classifications:
primary 62C10
62C20
secondary 62F15
62J07

Keywords:
Decision theory
Equivariance
Generalized Bayes estimation
Hierarchical model
Minimaxity
Multivariate linear model
Posterior mean
Quadratic loss
Shrinkage estimator

ABSTRACT

This paper addresses the problem of estimating the normal mean matrix in the case of unknown covariance matrix. This problem is solved by considering generalized Bayesian hierarchical models. The resulting generalized Bayes estimators with respect to an invariant quadratic loss function are shown to be matricial shrinkage equivariant estimators and the conditions for their minimaxity are given.

1. Introduction

The problem of estimating the normal mean matrix with known or unknown covariance matrices has been extensively studied from a decision-theoretic viewpoint in the literature, for instance, [1–10]. Their approaches to this problem include the empirical Bayes and the unbiased risk estimate methods. On the other hand, generalized Bayes procedures were recently applied in the case of known covariance. Berger et al. [11] gave general results on admissibility of certain hierarchical Bayes estimators and Tsukuma [12] proposed a Bayes minimax estimator. This paper addresses the problem of estimating the normal mean matrix in the case of unknown covariance and provides some generalized Bayes minimax estimators relative to an invariant quadratic loss function.

We begin with explaining the estimation problem considered in this paper. Let $\boldsymbol{X}$ be an $m \times p$ random matrix, where the row vectors, $\boldsymbol{x}_i$'s, are mutually independent and the i-th row vector $\boldsymbol{x}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\theta}_i$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Then $(\boldsymbol{x}_1^t, \dots, \boldsymbol{x}_m^t)^t$ follows multivariate normal distribution with mean vector $(\boldsymbol{\theta}_1^t, \dots, \boldsymbol{\theta}_m^t)^t$ and covariance matrix $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}$. Here $\boldsymbol{B}^t$ indicates the transpose of a vector (or matrix) $\boldsymbol{B}$, $\boldsymbol{I}_m$ is the identity matrix of order m and $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}$ indicates the Kronecker product of $\boldsymbol{I}_m$ and $\boldsymbol{\Sigma}$. Also, let $\boldsymbol{S}$ be a $p \times p$ random matrix having the Wishart distribution with n degrees of freedom and mean $n\boldsymbol{\Sigma}$. These models are written as

$$\boldsymbol{X} \sim \mathcal{N}_{m\times p}(\boldsymbol{\Theta}, \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}), \quad \boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}), \tag{1.1}$$

E-mail address: tsukuma@med.toho-u.ac.jp.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.04.009

where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^t$ and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)^t$. It is assumed that $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are unknown and that $\boldsymbol{X}$ and $\boldsymbol{S}$ are mutually independent. Note that the model (1.1) is a canonical form of a multivariate linear regression model. Our aim of this paper is to construct an eminent estimator of the mean matrix $\boldsymbol{\Theta}$ on the basis of $\boldsymbol{X}$ and $\boldsymbol{S}$ relative to invariant quadratic loss function

$$L(\delta, \boldsymbol{\Theta}; \boldsymbol{\Sigma}) = \operatorname{tr}(\delta - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\delta - \boldsymbol{\Theta})^t, \tag{1.2}$$

where $\delta = \delta(\boldsymbol{X}, \boldsymbol{S})$ is an estimator of $\boldsymbol{\Theta}$ and $\operatorname{tr}\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ denote, respectively, the trace and the inverse of a square matrix $\boldsymbol{A}$. Every estimator is evaluated by the risk function $E[L(\delta, \boldsymbol{\Theta}; \boldsymbol{\Sigma})]$, namely, the expected loss with respect to (1.1).

Let $\mathcal{O}_m$ be the set of orthogonal matrices of order m and let $\mathcal{P}$ be that of $p \times p$ nonsingular matrices. Denote $m \vee p = \max(m, p)$ and $m \wedge p = \min(m, p)$. Let $\boldsymbol{F} = \operatorname{diag}(f_1, \ldots, f_{m\wedge p})$ be a diagonal matrix based on ordered eigenvalues $f_1 \geq \cdots \geq f_{m\wedge p} \geq 0$, where, for $m < p$, $\boldsymbol{XS}^{-1}\boldsymbol{X}^t = \boldsymbol{RFR}^t$ with $\boldsymbol{R} \in \mathcal{O}_m$ and, for $m \geq p$, $\boldsymbol{Q}^t\boldsymbol{SQ} = \boldsymbol{I}_p$ and $\boldsymbol{Q}^t\boldsymbol{X}^t\boldsymbol{XQ} = \boldsymbol{F}$ with $\boldsymbol{Q} \in \mathcal{P}$. Konno [7–9] showed that for the group of transformations $\boldsymbol{X} \to \boldsymbol{OXV}$ and $\boldsymbol{S} \to \boldsymbol{V}^t\boldsymbol{SV}$ with any $\boldsymbol{O} \in \mathcal{O}_m$ and any $\boldsymbol{V} \in \mathcal{P}$, the class of equivariant estimators with matricial shrinkage factors is expressed by

$$\delta^{SH} = \begin{cases} (\boldsymbol{I}_m - \boldsymbol{RF}^{-1}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t)\boldsymbol{X} & \text{if } m < p, \\ \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{QF}^{-1}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^{-1}) & \text{if } m \geq p. \end{cases} \tag{1.3}$$

Here, $\boldsymbol{\Phi}(\boldsymbol{F}) = \operatorname{diag}(\phi_1(\boldsymbol{F}), \ldots, \phi_{m\wedge p}(\boldsymbol{F}))$ is certain $(m \wedge p) \times (m \wedge p)$ diagonal matrix with diagonal elements, $\phi_i(\boldsymbol{F})$'s, being functions of $\boldsymbol{F}$. The maximum likelihood estimator of $\boldsymbol{\Theta}$ is $\delta^{ML} = \boldsymbol{X}$, which has equivariance and is minimax with the constant risk mp. Konno [7–9] also showed that if

(i) for $i = 1, \ldots, m \wedge p$, $\phi_i(\boldsymbol{F})$ is nondecreasing in f_i,
(ii) $0 \leq \phi_{m\wedge p}(\boldsymbol{F}) \leq \phi_{m\wedge p-1}(\boldsymbol{F}) \leq \cdots \leq \phi_1(\boldsymbol{F}) \leq \frac{2(m\vee p - m\wedge p - 1)}{n + (2m-p)\wedge p + 1}$,

then δ^{SH} has smaller risk than δ^{ML} relative to the loss (1.2), namely, δ^{SH} is minimax.

In this paper, we consider generalized Bayes estimation of the mean matrix $\boldsymbol{\Theta}$ via hierarchical models. To specify the hierarchical prior distributions, we use the following notation. Let $\boldsymbol{0}_{c\times d}$ be the $c \times d$ zero matrix. Denote by $|\boldsymbol{A}|$ the determinant of a square matrix $\boldsymbol{A}$. For a positive definite matrix $\boldsymbol{B}$, let $\boldsymbol{B}^{1/2}$ be a symmetric matrix such that $\boldsymbol{B} = \boldsymbol{B}^{1/2}\boldsymbol{B}^{1/2}$ and $\boldsymbol{B}^{-1/2}$ stands for the inverse matrix of $\boldsymbol{B}^{1/2}$. If $\boldsymbol{A} - \boldsymbol{B}$ is positive definite for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, then we write $\boldsymbol{A} > \boldsymbol{B}$ or $\boldsymbol{B} < \boldsymbol{A}$. Also, let $I(\cdot)$ be the indicator function. In the case where $m < p$ we consider hierarchical Bayes model whose first stage prior of $\boldsymbol{\Theta}$ is

$$\boldsymbol{\Theta}|\boldsymbol{\Omega}, \boldsymbol{\Sigma}^{-1} \sim \mathcal{N}_{m\times p}(\boldsymbol{0}_{m\times p}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}) \tag{1.4}$$

and second stage priors of $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}^{-1}$ have densities proportional to, respectively,

$$\pi(\boldsymbol{\Omega}) \propto |\boldsymbol{I}_m + \boldsymbol{\Omega}|^{-a/2-m} I(\boldsymbol{\Omega} > \boldsymbol{0}_{m\times m}), \tag{1.5}$$

$$\pi(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{(b-1)/2} I(\boldsymbol{\Sigma}^{-1} > \boldsymbol{0}_{p\times p}). \tag{1.6}$$

For the case of $m \geq p$ we utilize the following hierarchical priors:

$$\boldsymbol{\Theta}|\boldsymbol{\Xi} \sim \mathcal{N}_{m\times p}(\boldsymbol{0}_{m\times p}, \boldsymbol{I}_m \otimes \boldsymbol{\Xi}), \tag{1.7}$$

$$\pi(\boldsymbol{\Xi}|\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{I}_p + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Xi}\boldsymbol{\Sigma}^{-1/2}|^{-a/2-p} I(\boldsymbol{\Xi} > \boldsymbol{0}_{p\times p}), \tag{1.8}$$

$$\pi(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{(b+p)/2} I(\boldsymbol{\Sigma}^{-1} > \boldsymbol{0}_{p\times p}). \tag{1.9}$$

These prior distributions are regarded as certain extensions of Lin and Tsai [13] for generalized Bayes minimax estimation of the normal mean vector.

The (generalized) Bayes estimator is usually defined as the one which minimizes the posterior expected loss. Our resulting generalized Bayes estimators against the above hierarchical priors with respect to the quadratic loss (1.2) are given by

$$\delta^{GB} = \begin{cases} E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}]\left\{E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Sigma}^{-1}]\right\}^{-1} & \text{if } m < p, \\ E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Xi},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}]\left\{E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Xi},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Sigma}^{-1}]\right\}^{-1} & \text{if } m \geq p, \end{cases}$$

where $E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}$ and $E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Xi},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}$ denote the expectations associated with the posterior distributions of $(\boldsymbol{\Theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}^{-1})$ and $(\boldsymbol{\Theta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}^{-1})$, respectively.

This paper concerns minimaxity of δ^{GB}. If δ^{GB} belong to the class (1.3) then Konno [7–9]' results given above enable us to evaluate the risk functions of δ^{GB}. Indeed, Section 2 gives that the generalized Bayes estimators δ^{GB} have equivariance with matricial shrinkage factors. Section 3 provides the conditions for minimaxity of δ^{GB} and shows that they are given as follows: (i) $a + m\vee p > 0$, (ii) $n - m\wedge p + b - a + 2 > 0$ and (iii) $(a + m + p - 1)/(n + b - a + 1) \leq 2(m\vee p - m\wedge p - 1)/\{n + (2m-p)\wedge p + 1\}$.

As the other Bayesian solutions, we may employ the posterior means against the above hierarchical priors $(\boldsymbol{\Theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}^{-1})$ and $(\boldsymbol{\Theta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}^{-1})$,

$$\delta^{PM} = \begin{cases} E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Theta}] & \text{if } m < p, \\ E_{\pi(\boldsymbol{\Theta},\boldsymbol{\Xi},\boldsymbol{\Sigma}^{-1}|\boldsymbol{X},\boldsymbol{S})}[\boldsymbol{\Theta}] & \text{if } m \geq p. \end{cases}$$

In Remarks of Sections 2 and 3, we state equivariance of δ^{PM} and the conditions for their minimaxity, which are extended results of Lin and Tsai [13]. Section 4 gives the concluding remarks of this paper and an open problem in this research area.