Journal of Multivariate Analysis 124 (2014) 333–344

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Minimax covariance estimation using commutator subgroup of lower triangular matrices

Hisayuki Tsukuma

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

ARTICLE INFO

Article history:
Received 14 October 2012
Available online 2 December 2013

AMS 2010 subject classifications:
primary 62C20
62F10
secondary 62H12

Keywords:
Commutator subgroup
Covariance matrix
Least favorable prior
Statistical decision theory
Stein's loss
Wishart distribution

ABSTRACT

This paper deals with the problem of estimating the normal covariance matrix relative to the Stein loss. The main interest concerns a new class of estimators which are invariant under a commutator subgroup of lower triangular matrices. The minimaxity of a James–Stein type invariant estimator under the subgroup is shown by means of a least favorable sequence of prior distributions. The class yields improved estimators on the James–Stein type invariant and minimax estimator.

1. Introduction

Let $\boldsymbol{S}$ be a $p \times p$ random Wishart matrix with n degrees of freedom and mean $n\boldsymbol{\Sigma}$. Suppose that $n \geq p$ and $\boldsymbol{\Sigma}$ is positive definite. In this paper, the matrix $\boldsymbol{\Sigma}$ is referred to as covariance matrix of a multivariate normal distribution. Consider then the problem of estimating the covariance matrix $\boldsymbol{\Sigma}$ relative to the Stein loss

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}| - p, \tag{1.1}$$

where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\Sigma}$ based only on $\boldsymbol{S}$. The quality of estimators is measured by the corresponding risk function $R(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = E[L(\boldsymbol{\delta}, \boldsymbol{\Sigma})]$.

Denote by $\mathcal{T}^+$ the group of $p \times p$ lower triangular matrices with positive diagonal entries such that the group $\mathcal{T}^+$ acts on $\boldsymbol{S}$ and $\boldsymbol{\delta}$ as follows: $\boldsymbol{S} \to \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t$ and $\boldsymbol{\delta} \to \boldsymbol{A}\boldsymbol{\delta}\boldsymbol{A}^t$ for any $\boldsymbol{A} \in \mathcal{T}^+$. Let $\mathcal{D}^+$ be the group of diagonal matrices of order p with positive diagonal entries. The Cholesky decomposition of $\boldsymbol{S}$ stands for $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^t$, where $\boldsymbol{T} \in \mathcal{T}^+$ and $\boldsymbol{T}$ is uniquely determined. Stein [13] and James and Stein [9] considered invariant estimators under $\mathcal{T}^+$ and showed minimaxity of the best invariant estimator relative to the Stein loss (1.1). It is noted that the best invariant and minimax estimator is given by $\boldsymbol{\delta}^{JS} = \boldsymbol{T}\boldsymbol{D}^{BI}\boldsymbol{T}^t$, where $\boldsymbol{D}^{BI} \in \mathcal{D}^+$ and the ith diagonal entry of $\boldsymbol{D}^{BI}$ is given by $(n + p - 2i + 1)^{-1}$. See also [4] for derivation of $\boldsymbol{\delta}^{JS}$.

The discovery of the best invariant and minimax estimator $\boldsymbol{\delta}^{JS}$ immediately results in the unbiased and maximum likelihood estimator $\boldsymbol{S}/n$ being not minimax. Since this surprising result, many researchers have made efforts to obtain minimax estimators of the normal covariance matrix. For instance, see [3,14,15,2,12,7] and others. Recently, Tsukuma and Kubokawa [17] treated a Bayesian approach to minimaxity of the best equivariant estimators under the group $\mathcal{T}^+$ in estimation of restricted and non-restricted scale parameter matrices in a more general model.

E-mail addresses: tsukuma@med.toho-u.ac.jp, tsuku@kk.iij4u.or.jp.

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.11.007

A significant property of the group $\mathcal{T}^+$ is solvability. This indeed plays an important role in proof for minimaxity of $\boldsymbol{\delta}^{JS}$. The interesting connectivity of solvability to the proof is revealed as the Hunt–Stein theorem. For details of the Hunt–Stein theorem, see [10,1]. Meanwhile, the solvability of the group $\mathcal{T}^+$ implies that $\mathcal{T}^+$ includes a commutator subgroup (also called derived subgroup). In this paper, we direct our attention to the commutator subgroup of $\mathcal{T}^+$ and deal with invariant estimators under the subgroup.

Section 2 provides an explicit formula for the class of invariant estimators under the commutator subgroup of $\mathcal{T}^+$ and gives a James–Stein type best invariant estimator among a subclass of the class. Section 3 shows that the James–Stein type invariant estimator is minimax relative to the Stein loss (1.1). The proof of minimaxity is based on a Bayesian approach with a least favorable sequence of prior distributions, but not on the Hunt–Stein theorem. In Section 4, we obtain an improved estimator on the James–Stein type invariant and minimax estimator, and Section 5 gives numerical results to investigate the risk performance of minimax estimators. Section 6 states some concluding remarks with open problems.

2. Invariant estimators under a commutator subgroup

For two elements $\boldsymbol{A}$ and $\boldsymbol{B}$ of the group $\mathcal{T}^+$, the commutator of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined by $\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$. The commutator subgroup of $\mathcal{T}^+$ is generated by all the commutators of $\mathcal{T}^+$. Denote the commutator subgroup of $\mathcal{T}^+$ by $\mathcal{T}^{(1)}$. It is here noted that $\mathcal{T}^{(1)}$ consists of $p\times p$ lower triangular matrices with ones on the diagonal.

Let $\boldsymbol{S}=\boldsymbol{T}_1\boldsymbol{T}_0\boldsymbol{T}_1^t$, where $\boldsymbol{T}_0$ and $\boldsymbol{T}_1$ are unique elements of $\mathcal{D}^+$ and $\mathcal{T}^{(1)}$, respectively. The decomposition is often referred to as the LDL^t decomposition (factorization). For the existence and the uniqueness of the decomposition, see Golub and Van Loan [6, Section 4.1], who also provide an algorithm of the decomposition.

Define the group action on $\boldsymbol{S}$ and $\boldsymbol{\delta}$ by $\mathcal{T}^{(1)}$ as $\boldsymbol{S}\to\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t$ and $\boldsymbol{\delta}\to\boldsymbol{A}\boldsymbol{\delta}\boldsymbol{A}^t$ for any $\boldsymbol{A}\in\mathcal{T}^{(1)}$. Then invariant estimators under the commutator subgroup $\mathcal{T}^{(1)}$ are expressed as $\boldsymbol{T}_1\boldsymbol{G}(\boldsymbol{T}_0)\boldsymbol{T}_1^t$, where $\boldsymbol{G}(\boldsymbol{T}_0)$ is a symmetric matrix whose entries are functions of $\boldsymbol{T}_0$. Because it is very hard to handle the invariant estimators described above, we hereafter restrict ourselves to the following class:

$$\boldsymbol{\delta}^I=\boldsymbol{\delta}^I(\boldsymbol{T}_0,\boldsymbol{T}_1)=\boldsymbol{T}_1\boldsymbol{\Phi}(\boldsymbol{T}_0)\boldsymbol{T}_1^t, \tag{2.1}$$

where $\boldsymbol{\Phi}(\boldsymbol{T}_0)$ lies in $\mathcal{D}^+$ and each diagonal entry of $\boldsymbol{\Phi}(\boldsymbol{T}_0)$ is a function of $\boldsymbol{T}_0$.

The joint probability density function (p.d.f.) of $(\boldsymbol{T}_0,\boldsymbol{T}_1)$ is derived from the p.d.f. of $\boldsymbol{S}$, which is written as

$$f(\boldsymbol{S}|\boldsymbol{\Sigma}^{-1})=c_{n,p}|\boldsymbol{\Sigma}^{-1}|^{n/2}|\boldsymbol{S}|^{(n-p-1)/2}e^{-(1/2)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{S}},$$

where $c_{n,p}$ is a normalizing constant. The Jacobian of transformation from $\boldsymbol{S}$ to $(\boldsymbol{T}_0,\boldsymbol{T}_1)$ is $J[\boldsymbol{S}\to(\boldsymbol{T}_0,\boldsymbol{T}_1)]=\prod_{i=1}^p t_i^{p-i}$ with $\boldsymbol{T}_0=\mathrm{diag}(t_1,\dots,t_p)$. Denote by $\boldsymbol{\Sigma}=\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t$ the unique decomposition of $\boldsymbol{\Sigma}$, where $\boldsymbol{\Gamma}_1\in\mathcal{T}^{(1)}$ and $\boldsymbol{\Gamma}_0=\mathrm{diag}(\gamma_1,\dots,\gamma_p)\in\mathcal{D}^+$. It follows that $|\boldsymbol{S}|=\prod_{i=1}^p t_i$ and $|\boldsymbol{\Sigma}^{-1}|=\prod_{i=1}^p\gamma_i^{-1}$. Hence the joint p.d.f. of $(\boldsymbol{T}_0,\boldsymbol{T}_1)$ is given by

$$f(\boldsymbol{T}_0,\boldsymbol{T}_1|\boldsymbol{\Gamma}_0^{-1},\boldsymbol{\Gamma}_1^{-1})=c_{n,p}\left(\prod_{i=1}^p t_i^{(n+p-2i-1)/2}\right)\left(\prod_{i=1}^p\gamma_i^{-n/2}\right)e^{-(1/2)\mathrm{tr}\,(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t)^{-1}\boldsymbol{T}_1\boldsymbol{T}_0\boldsymbol{T}_1^t}. \tag{2.2}$$

It is worth noting that the unique decomposition $\boldsymbol{\Sigma}=\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t$ is characterized by Pourahmadi [11], who pointed out that the diagonal entries of $\boldsymbol{\Gamma}_0$ and the entries below unit diagonals of $\boldsymbol{\Gamma}_1^{-1}$ are interpreted as the innovation variances and the autoregressive coefficients, respectively, in statistical analysis for longitudinal data.

We next derive a James–Stein type best invariant estimator among a subclass of (2.1). As pointed out by Eaton [4], Theorem 6.5), the best invariant estimator is equivalent to a generalized Bayes estimator against a right invariant measure used as a prior distribution. The right invariant measures on $\mathcal{T}^{(1)}$ and $\mathcal{D}^+$ under scale transformations are expressed by $\mu(\mathrm{d}\boldsymbol{\Gamma}_1)=\mathrm{d}\boldsymbol{\Gamma}_1$ and $\mu(\mathrm{d}\boldsymbol{\Gamma}_0)=(\prod_{i=1}^p\gamma_i^{-1})\mathrm{d}\boldsymbol{\Gamma}_0$, respectively. Then, the best invariant estimator $\boldsymbol{\delta}^{BI}=\boldsymbol{\delta}^{BI}(\boldsymbol{T}_0,\boldsymbol{T}_1)$ with respect to the Stein loss (1.1) has the form:

Lemma 2.1.

$$\boldsymbol{\delta}^{BI}=\boldsymbol{\delta}^{BI}(\boldsymbol{T}_0,\boldsymbol{T}_1)=\boldsymbol{T}_1\boldsymbol{\Phi}^{BI}(\boldsymbol{T}_0)\boldsymbol{T}_1^t, \tag{2.3}$$

where $\boldsymbol{\Phi}^{BI}(\boldsymbol{T}_0)=\mathrm{diag}(d_1^{BI}t_1,\dots,d_p^{BI}t_p)$ *with* $d_i^{BI}=(n+p-2i+1)^{-1}$ *for* $i=1,\dots,p$.

Proof. The generalized Bayes estimator against prior $\mu(\mathrm{d}\boldsymbol{\Gamma}_0)\mu(\mathrm{d}\boldsymbol{\Gamma}_1)$ is expressed by

$$\begin{aligned}\boldsymbol{\delta}^{BI}&=\left\{\frac{\int_{\mathcal{D}^+\times\mathcal{T}^{(1)}}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t)^{-1}f(\boldsymbol{T}_0,\boldsymbol{T}_1|\boldsymbol{\Gamma}_0^{-1},\boldsymbol{\Gamma}_1^{-1})\mu(\mathrm{d}\boldsymbol{\Gamma}_1)\mu(\mathrm{d}\boldsymbol{\Gamma}_0)}{\int_{\mathcal{D}^+\times\mathcal{T}^{(1)}}f(\boldsymbol{T}_0,\boldsymbol{T}_1|\boldsymbol{\Gamma}_0^{-1},\boldsymbol{\Gamma}_1^{-1})\mu(\mathrm{d}\boldsymbol{\Gamma}_1)\mu(\mathrm{d}\boldsymbol{\Gamma}_0)}\right\}^{-1}\\&=\left\{\frac{\int_{\mathcal{D}^+\times\mathcal{T}^{(1)}}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t)^{-1}e^{-(1/2)\mathrm{tr}\,(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t)^{-1}\boldsymbol{T}_1\boldsymbol{T}_0\boldsymbol{T}_1^t}\left(\prod_{i=1}^p\gamma_i^{-n/2-1}\right)\mathrm{d}\boldsymbol{\Gamma}_0\mathrm{d}\boldsymbol{\Gamma}_1}{\int_{\mathcal{D}^+\times\mathcal{T}^{(1)}}e^{-(1/2)\mathrm{tr}\,(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_1^t)^{-1}\boldsymbol{T}_1\boldsymbol{T}_0\boldsymbol{T}_1^t}\left(\prod_{i=1}^p\gamma_i^{-n/2-1}\right)\mathrm{d}\boldsymbol{\Gamma}_0\mathrm{d}\boldsymbol{\Gamma}_1}\right\}^{-1}.\end{aligned}$$