Journal of Multivariate Analysis 140 (2015) 245–258

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Estimation of the mean vector in a singular multivariate normal distribution

Hisayuki Tsukuma [a,*], Tatsuya Kubokawa [b]

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

[b] *Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

ARTICLE INFO

Article history:
Received 24 April 2014
Available online 29 May 2015

AMS 2010 subject classifications:
primary 62C20
62J07
secondary 62C10

Keywords:
Empirical Bayes method
Generalized Bayes estimator
Inadmissibility
Minimaxity
Moore–Penrose inverse
Pseudo-Wishart distribution
Quadratic loss
Shrinkage estimator
Statistical decision theory

ABSTRACT

This paper addresses the problem of estimating the mean vector of a singular multivariate normal distribution with an unknown singular covariance matrix. The maximum likelihood estimator is shown to be minimax relative to a quadratic loss weighted by the Moore–Penrose inverse of the covariance matrix. An unbiased risk estimator relative to the weighted quadratic loss is provided for a Baranchik type class of shrinkage estimators. Based on the unbiased risk estimator, a sufficient condition for the minimaxity is expressed not only as a differential inequality, but also as an integral inequality. Also, generalized Bayes minimax estimators are established by using an interesting structure of singular multivariate normal distribution.

1. Introduction

Statistical inference with the determinant and the inverse of sample covariance matrix requires nonsingularity of the sample covariance matrix. However in practical cases of data analysis, the nonsingularity is not always satisfied. The singularity occurs for many reasons, but in general such singularity is very hard to handle. This paper treats a singular multivariate normal model, which yields a singular sample covariance matrix, and aims to provide a series of decision-theoretic results in estimation of the mean vector.

The singular multivariate normal distribution model and the related topics have been studied for a long time in the literature. For the density function, see [7,12,15]. Khatri [7] and Rao [12] derived the maximum likelihood estimators for the mean vector and the singular covariance matrix. Srivastava [13] and Díaz-García, *et al.* [3] studied central and noncentral pseudo-Wishart distributions which have been used for developing distribution theories in the problems of testing hypotheses. However, little is known about a decision-theoretic approach to estimation in the singular model.

To specify the singular model addressed in this paper, let $\boldsymbol{X}$ and $\boldsymbol{Y}_i$ ($i = 1, \ldots, n$) be p-dimensional random vectors having the stochastic representations

* Corresponding author.
E-mail addresses: tsukuma@med.toho-u.ac.jp (H. Tsukuma), tatsuya@e.u-tokyo.ac.jp (T. Kubokawa).

http://dx.doi.org/10.1016/j.jmva.2015.05.016
0047-259X/

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{Z}_0, \\ \boldsymbol{Y}_i &= \boldsymbol{B}\boldsymbol{Z}_i, \quad i = 1, \ldots, n, \end{aligned} \tag{1.1}$$

where $\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are mutually and independently distributed as $\mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{I}_r)$, and $\boldsymbol{\theta}$ and $\boldsymbol{B}$ are, respectively, a p-dimensional vector and a $p \times r$ matrix of unknown parameters. Then we write $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y}_i \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ $(i = 1, \ldots, n)$, where $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^t$. Assume that

$$r \le \min(n, p),$$

and $\boldsymbol{B}$ is of full column rank, namely $\boldsymbol{\Sigma}$ is a positive semi-definite matrix of rank r. In the case when $r < p$, technically speaking, $\mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is called the singular multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and singular covariance $\boldsymbol{\Sigma}$. For the definition of the singular multivariate normal distribution, see [7], [12, Chapter 8] and [15, p. 43].

Denote by $\boldsymbol{\Sigma}^+$ the Moore–Penrose inverse of $\boldsymbol{\Sigma}$. Consider the problem of estimating the mean vector $\boldsymbol{\theta}$ relative to quadratic loss weighted by $\boldsymbol{\Sigma}^+$,

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = (\boldsymbol{\delta} - \boldsymbol{\theta})^t \boldsymbol{\Sigma}^+ (\boldsymbol{\delta} - \boldsymbol{\theta}), \tag{1.2}$$

where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\theta}$ based on $\boldsymbol{X}$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^t$. The accuracy of estimators is compared by the risk function $R(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) = E[L(\boldsymbol{\delta}, \boldsymbol{\theta}|\boldsymbol{\Sigma})]$, where the expectation is taken with respect to (1.1).

A natural estimator of $\boldsymbol{\theta}$ is the unbiased estimator $\boldsymbol{\delta}^{UB} = \boldsymbol{X}$, which is also the maximum likelihood estimator as pointed out by Khatri [7, p. 276] and Rao [12, p. 532]. This paper considers improvement on $\boldsymbol{\delta}^{UB}$ via the Baranchik [1] type class of shrinkage estimators

$$\boldsymbol{\delta}^{SH} = \Big(1 - \frac{\phi(F)}{F}\Big)\boldsymbol{X}, \quad F = \boldsymbol{X}^t \boldsymbol{S}^+ \boldsymbol{X},$$

where $\phi(F)$ is a bounded and differentiable function of F.

It is worth noting that, instead of F in $\boldsymbol{\delta}^{SH}$, we may use $F^- = \boldsymbol{X}^t \boldsymbol{S}^- \boldsymbol{X}$, where $\boldsymbol{S}^-$ is a generalized inverse of $\boldsymbol{S}$. Since the generalized inverse is not unique, it may be troublesome to consider which we employ as the generalized inverse. On the other hand, the Moore–Penrose inverse is unique and it is easy to discuss its distributional property. See [14] for interesting discussion on the Hotelling type T-square tests with the Moore–Penrose and the generalized inverses in high dimension.

The rest of this paper is organized as follows. In Section 2, we introduce the definition of the Moore–Penrose inverse and its useful properties. We then set up a decision-theoretic framework for estimating $\boldsymbol{\theta}$ and derive some properties of estimators and their risk functions which are specific to the singular model. The key tool for their derivations is the equality

$$\boldsymbol{S}\boldsymbol{S}^+ = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^+,$$

which holds with probability one, where $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y}$ and $\boldsymbol{S}^+$ is the Moore–Penrose inverse of $\boldsymbol{S}$. In Section 2, we also prove the minimaxity of $\boldsymbol{\delta}^{UB}$. In Section 3, we obtain sufficient conditions for the minimaxity of $\boldsymbol{\delta}^{SH}$. These conditions are given not only by a differential inequality, but also by an integral inequality. In Section 4, an empirical Bayes motivation is given for the James–Stein [6] type shrinkage estimator and its positive part estimator. Also, Section 4 suggests a hierarchical prior in the singular model and shows that the resulting generalized Bayes estimators are minimax. Section 5 provides some remarks on related topics.

2. Estimation in the singular normal model

2.1. The Moore–Penrose inverse and its useful properties

We begin by introducing the following notations which will be used through the paper. Let $\mathcal{O}(r)$ be the group of orthogonal matrices of order r. For $p \ge r$, the Stiefel manifold is denoted by $\mathcal{V}_{p,r} = \{\boldsymbol{A} \in \mathbb{R}^{p\times r} : \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I}_r\}$. It is noted that $\mathcal{V}_{r,r} = \mathcal{O}(r)$. Let $\mathcal{D}_r$ be a set of $r \times r$ diagonal matrices whose diagonal elements $d_1, \ldots, d_r$ satisfy $d_1 > \cdots > d_r > 0$.

As an inverse matrix of a singular covariance matrix, we use the Moore–Penrose inverse matrix, which is defined as follows:

Definition 2.1. For a matrix $\boldsymbol{A}$, there exists a matrix $\boldsymbol{A}^+$ such that (i) $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$, (ii) $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$, (iii) $(\boldsymbol{A}\boldsymbol{A}^+)^t = \boldsymbol{A}\boldsymbol{A}^+$ and (iv) $(\boldsymbol{A}^+\boldsymbol{A})^t = \boldsymbol{A}^+\boldsymbol{A}$. Then $\boldsymbol{A}^+$ is called the Moore–Penrose inverse of $\boldsymbol{A}$.

The following basic properties and results on the Moore–Penrose inverse matrix are useful for investigating properties of shrinkage estimators. Lemmas 2.1 and 2.2 are due to Harville [5, Chapter 20].

Lemma 2.1. *The Moore–Penrose inverse $\boldsymbol{A}^+$ has the following properties:*

(1) *$\boldsymbol{A}^+$ uniquely exists;*
(2) *$(\boldsymbol{A}^+)^t = (\boldsymbol{A}^t)^+$;*
(3) *$\boldsymbol{A}^+ = \boldsymbol{A}^{-1}$ for a nonsingular matrix $\boldsymbol{A}$.*