Journal of Multivariate Analysis 139 (2015) 312–328

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A unified approach to estimating a normal mean matrix in high and low dimensions

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 29 March 2014
Available online 16 April 2015

AMS 2010 subject classifications:
primary 62C10
secondary 62F10

Keywords:
Efron–Morris estimator
Empirical Bayes procedure
High dimension
Invariant loss
Matrix mean
Moore–Penrose inverse
Shrinkage estimator
Statistical decision theory

ABSTRACT

This paper addresses the problem of estimating the normal mean matrix with an unknown covariance matrix. Motivated by an empirical Bayes method, we suggest a unified form of the Efron–Morris type estimators based on the Moore–Penrose inverse. This form not only can be defined for any dimension and any sample size, but also can contain the Efron–Morris type or Baranchik type estimators suggested so far in the literature. Also, the unified form suggests a general class of shrinkage estimators. For shrinkage estimators within the general class, a unified expression of unbiased estimators of the risk functions is derived regardless of the dimension of covariance matrix and the size of the mean matrix. An analytical dominance result is provided for a positive-part rule of the shrinkage estimators.

1. Introduction

Statistical inference with high dimension has received much attention in recent years, because statistical analysis of high-dimensional data has been requested in many research areas such as genomics, remote sensing, telecommunication, atmospheric science, financial engineering, and others. Such high-dimensional data are generally hard to handle, and ordinary or traditional methods are frequently inapplicable. This has inspired statisticians to develop new research areas in high dimension from both theoretical and practical aspects. Most interests have been in development of efficient algorithm for statistical inference and in derivation of asymptotic properties with the dimension going to infinity. From a decision-theoretic point of view, however, there does not exist much literature in high-dimensional problems except for Chélat and Wells [2], who established the inadmissibility of the maximum likelihood estimator (MLE) for a large dimensional and small sample normal model. In this paper, we extend their result to the framework of estimating a mean matrix and we establish a unified theory for improvement on the MLE in both cases of high and low dimensions.

To explain the subjects addressed here, we begin by describing the canonical model and the estimation problem. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_m)^t$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n)^t$ be, respectively, $m \times p$ and $n \times p$ random matrices, where $\boldsymbol{X}_i$'s and $\boldsymbol{Y}_i$'s are mutually and independently distributed as

$$\begin{aligned} \boldsymbol{X}_i &\sim \mathcal{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}), \quad i = 1, \dots, m, \\ \boldsymbol{Y}_j &\sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma}), \quad j = 1, \dots, n. \end{aligned} \tag{1.1}$$

* 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.2015.04.003

Suppose that $\boldsymbol{\theta}_i$'s are unknown mean vectors and that $\boldsymbol{\Sigma}$ is an unknown positive definite matrix. It is noted that the model (1.1) is a canonical form of a multivariate linear regression model although the details are omitted here. The problem we consider in this paper is the estimation of the mean matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)^t$ relative to the invariant quadratic loss

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

where δ is an estimator made from $\boldsymbol{X}$ and $\boldsymbol{S}$.

The MLE of $\boldsymbol{\Theta}$ is $\delta^{ML} = \boldsymbol{X}$, which is a minimax estimator with the constant risk mp. When $n \ge p$, it is known that δ^{ML} is improved on by the Efron–Morris [3] type estimator

$$\delta^{EMK} = \begin{cases} \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1}\boldsymbol{X} & \text{if } p \ge m, \\ \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S} & \text{if } m > p, \end{cases}$$

where $\boldsymbol{S} = \boldsymbol{Y}^t\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{Y}_i\boldsymbol{Y}_i^t$, and c is a suitable constant. Konno [5,6] derived conditions on c for the improvement. When $p > n$, however, this estimator is not available because the inverse $\boldsymbol{S}^{-1}$ does not exist. A possible alternative is the Moore–Penrose inverse $\boldsymbol{S}^+$ which will be defined in the beginning of Section 2. In the case of $m = 1$, Chélat and Wells [2] suggested the shrinkage estimator

$$\delta^{CW} = \boldsymbol{X} - \frac{c}{\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t}\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+$$

and provided a condition on c for δ^{CW} to dominate δ^{ML}.

An interesting issue is how to extend the Chételat–Wells estimator δ^{CW} to the framework of estimation of the mean matrix $\boldsymbol{\Theta}$. Especially, the Efron–Morris type estimators seem to take various variants which depend on orderings among m, p and n. One of interesting results provided in this paper is that we can develop a unified form for the Efron–Morris type estimators, given by

$$\delta^{EM} = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{S}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{S}\boldsymbol{S}^+.$$

As explained in Section 2, this estimator can be defined for all the positive integers of m, p and n as well as this expression includes δ^{EMK} and δ^{CW} as special cases. Also this expression suggests us to consider a general class of shrinkage estimators in Section 3. In this paper, we derive a unified expression of an unbiased estimator of the risk function for shrinkage estimators within the general class.

The paper is organized as follows. In Section 2, we introduce shrinkage estimators of $\boldsymbol{\Theta}$ based on a motivation from an empirical Bayes method, and we provide the unified form of the Efron–Morris type estimator. This expression not only contains the Efron–Morris or Baranchik type estimators suggested so far in the literature, but also provide various forms corresponding to ordering of m, p and n. In Section 3, we consider a general class of shrinkage estimators. A unified expression is developed in Section 4 for the risk functions of the general shrinkage estimators. It is noted that the unified expression gives an unbiased estimator of the risk difference. As specific examples of shrinkage estimators, we treat the modified Efron–Morris type estimators and the modified Stein type estimators, and we get conditions for their improvement from the unified expression. Section 5 provides analytical and numerical dominance results that positive-part estimators improve the corresponding shrinkage estimators. Some technical proofs are given in Section 6.

2. A Bayesian motivation

We begin by describing basic and useful properties of the Moore–Penrose inverse. For any matrix $\boldsymbol{A}$, the Moore–Penrose inverse of $\boldsymbol{A}$ is written by $\boldsymbol{A}^+$ if $\boldsymbol{A}^+$ satisfies (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) $(\boldsymbol{A}^+\boldsymbol{A})^t = \boldsymbol{A}^+\boldsymbol{A}$. 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 $\boldsymbol{A}$.

Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be $r \times p$ matrices of full row rank. We then have (1) $\boldsymbol{B}^+ = \boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{B}^t)^{-1}$, (2) $\boldsymbol{B}\boldsymbol{B}^+ = \boldsymbol{I}_r$, (3) $\boldsymbol{B}^+\boldsymbol{B}$ is idempotent, (4) $(\boldsymbol{B}^t\boldsymbol{C})^+ = \boldsymbol{C}^+(\boldsymbol{B}^t)^+ = \boldsymbol{C}^t(\boldsymbol{C}\boldsymbol{C}^t)^{-1}(\boldsymbol{B}\boldsymbol{B}^t)^{-1}\boldsymbol{B}$. Further, for an $r \times r$ nonsingular matrix $\boldsymbol{A}$ and an $r \times q$ matrix $\boldsymbol{B}$ of full row rank, we can easily show that $(\boldsymbol{B}^t\boldsymbol{A}\boldsymbol{B})^+ = \boldsymbol{B}^+\boldsymbol{A}^{-1}(\boldsymbol{B}^t)^+$.

Based on the properties of the Moore–Penrose inverse, we give a unified form of empirical Bayes estimators. Using a similar argument as in Tsukuma and Kubokawa [14], we can show that in the case of known $\boldsymbol{\Sigma}$, the empirical Bayes estimator of $\boldsymbol{\Theta}$ is given by

$$\delta^{B} = \begin{cases} \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^{-1}\boldsymbol{X} & \text{for } p \ge m, \\ \boldsymbol{X} - c\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\Sigma} & \text{for } m > p, \end{cases}$$

for a suitable constant c. Here it is observed that, for $m > p$,

$$(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+ = (\boldsymbol{X}^t)^+\boldsymbol{\Sigma}\boldsymbol{X}^+ = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\Sigma}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t,$$

which yields that $(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\Sigma}$. Hence, both cases $p \ge m$ and $m > p$ for the empirical Bayes estimator δ^B can be unified by

$$\delta^{B} = \boldsymbol{X} - c(\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^t)^+\boldsymbol{X}.$$