Journal of Multivariate Analysis 102 (2011) 801–815

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Admissible estimator of the eigenvalues of the variance–covariance matrix for multivariate normal distributions

Yo Sheena [a,*], Akimichi Takemura [b]

[a] Department of Economics, Shinshu University, Japan

[b] Graduate School of Information Science and Technology, University of Tokyo, Japan

ARTICLE INFO

Article history:
Received 7 September 2009
Available online 6 January 2011

AMS 2000 subject classifications:
primary 62C15
secondary 62F10

Keywords:
Covariance matrix
Wishart distribution
Squared error loss
Karlin's method

ABSTRACT

An admissible estimator of the eigenvalues of the variance–covariance matrix is given for multivariate normal distributions with respect to the scale-invariant squared error loss.

1. Introduction

The variance–covariance matrix of a multivariate normal distribution is usually estimated by the sample variance–covariance matrix, which is distributed as a Wishart distribution. Let $\boldsymbol{S}$ be distributed according to Wishart distribution $\boldsymbol{W}_p(\nu, \boldsymbol{\Sigma})$, where $p(\geq 2)$ is the dimension, $\nu(\geq p)$ is the degree of freedom, and $\boldsymbol{\Sigma}$ is the variance–covariance matrix of the original multivariate normal distribution.

In many situations of multivariate analysis, such as principle component analysis and canonical correlation analysis, we need to estimate the eigenvalues of $\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}$ itself. Also, many test statistics in multivariate analysis have distributions determined solely by the eigenvalues of $\boldsymbol{\Sigma}$ because of their invariance property under some natural transformations.

For the estimation of the eigenvalues of $\boldsymbol{\Sigma}$, the corresponding sample eigenvalues of $\boldsymbol{S}$ are usually used, but their distribution is quite complicated and makes it difficult to obtain mathematically clear results. Especially in a decision theoretic approach we encounter a difficulty since we essentially need the calculation of the risk (the expectation of a loss) with respect to the distribution of the eigenvalues for finite degrees of freedom ν. Mainly because of this difficulty, there exist only a few papers which directly deal with the estimation of the eigenvalues from the standpoint of the decision theory. Dey [1] and Jin [3] derive estimators which dominate the traditional estimators under the (non-scale-invariant) quadratic loss function. In view of the decision theory, one of the important tasks is to derive an admissible estimator, but it has been an unsolved problem so far. The aim of this paper is the derivation of an admissible estimator. For the proof of admissibility, we adopted the method of Ghosh and Singh [2],in which they proved the admissibility of an estimator for the reciprocal of the scale parameter of Gamma distributions using "Karlin's method" [4].

Here we formally state the framework. Let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ denote the eigenvalues of $\boldsymbol{\Sigma}$, while $l_1 \geq \cdots \geq l_p > 0$ are the eigenvalues of $\boldsymbol{S}$. As is well known, the distribution of $\boldsymbol{l} = (l_1, \ldots, l_p)$ depends only on $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$. For an

* Corresponding author.
E-mail address: sheena@shinshu-u.ac.jp (Y. Sheena).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.12.007

estimator

$$\boldsymbol{\psi}(\boldsymbol{l}) = (\psi_1(\boldsymbol{l}), \ldots, \psi_p(\boldsymbol{l})),$$

we measure the loss by the scale-invariant squared error loss function

$$\sum_{i=1}^{p}(\psi_i(\boldsymbol{l}) - \lambda_i)^2/\lambda_i^2 = \sum_{i=1}^{p}(\psi_i(\boldsymbol{l})/\lambda_i - 1)^2. \tag{1}$$

2. Main result

Before stating the main result as a theorem, we introduce some notation. For a vector $\boldsymbol{x} = (x_1, \ldots, x_p)$ and a set of powers $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ the monomial $x_1^{\alpha_1} \ldots x_p^{\alpha_p}$ is denoted by $\boldsymbol{x}^{\boldsymbol{\alpha}}$. If $\alpha = \alpha_1 = \cdots = \alpha_p$ is common, we denote the monomial by $\boldsymbol{x}^{\alpha}$. Let $\boldsymbol{H} = (h_{ij})$ denote a p-dimensional orthogonal matrix. The group of p dimensional orthogonal matrices is denoted by $\mathcal{O}(p)$ and μ is the invariant probability measure on $\mathcal{O}(p)$. Since we mainly work with the reciprocal of the population eigenvalue, $t_j = \lambda_j^{-1}$ $(j = 1, \ldots, p)$, more often than λ_j itself, we define the following notation for convenience.

$$\mathfrak{T}_a^b = \{\boldsymbol{t} = (t_1, \ldots, t_p) \mid (0 \le)\ a < t_1 \le \cdots \le t_p < b\ (\le \infty)\},$$

$$G(\boldsymbol{l}) = \boldsymbol{l}^{(\nu-p-1)/2} \prod_{i<j}(l_i - l_j),$$

$$F(\boldsymbol{t}|\boldsymbol{l}) = \int_{\mathcal{O}(p)} \exp\left(-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} t_i l_j h_{ij}^2\right) \mathrm{d}\mu(\boldsymbol{H}),$$

$$\partial_i F(\boldsymbol{t}|\boldsymbol{l}) = \frac{\partial F(\boldsymbol{t}|\boldsymbol{l})}{\partial t_i}, \quad i = 1, \ldots, p.$$

The density function $f(\boldsymbol{l}|\boldsymbol{t})$ of $\boldsymbol{l}$ is given by

$$f(\boldsymbol{l}|\boldsymbol{t}) = K\boldsymbol{t}^{\nu/2} G(\boldsymbol{l}) F(\boldsymbol{t}|\boldsymbol{l}), \tag{2}$$

where K is a constant (not depending on $\boldsymbol{l}$ and $\boldsymbol{t}$). Our main result is given as follows.

Theorem 1. *For $1 \le i \le p$, let*

$$\psi_i^*(\boldsymbol{l}) = -\left(\frac{\nu}{2} + 1\right)^{-1} \frac{\int_{\mathfrak{T}_0^\infty} \partial_i F(\boldsymbol{t}|\boldsymbol{l}) \boldsymbol{t}^{\nu/2-1} t_i^2 \mathrm{d}\boldsymbol{t}}{\int_{\mathfrak{T}_0^\infty} F(\boldsymbol{t}|\boldsymbol{l}) \boldsymbol{t}^{\nu/2-1} t_i^2 \mathrm{d}\boldsymbol{t}}.$$

The estimator $\boldsymbol{\psi}^*(\boldsymbol{l}) = (\psi_1^*(\boldsymbol{l}), \ldots, \psi_p^*(\boldsymbol{l}))$ is admissible with respect to the loss function (1).

Remark. From the argument on p. 201 of Stein [7], we see that $\boldsymbol{\psi}^*$ is admissible in the whole class of estimators of population eigenvalues, including estimators which also use the sample eigenvectors.

Proof of this theorem is given in Section 4.

Notice that $\psi_i^*(\boldsymbol{l})$ $(1 \le i \le p)$ can be rewritten as

$$\psi_i^*(\boldsymbol{l}) = \sum_{j=1}^{p} \tau_{ij}(\boldsymbol{l}) l_j, \tag{3}$$

where

$$(\nu+2)\tau_{ij}(\boldsymbol{l}) = \frac{\int_{\mathfrak{T}_0^\infty} \int_{\mathcal{O}(p)} h_{ij}^2 \exp\left(-\frac{1}{2}\sum_{s=1}^{p}\sum_{k=1}^{p} t_s l_k h_{sk}^2\right) \boldsymbol{t}^{\nu/2-1} t_i^2 \mathrm{d}\mu(\boldsymbol{H})\mathrm{d}\boldsymbol{t}}{\int_{\mathfrak{T}_0^\infty} \int_{\mathcal{O}(p)} \exp\left(-\frac{1}{2}\sum_{s=1}^{p}\sum_{k=1}^{p} t_s l_k h_{sk}^2\right) \boldsymbol{t}^{\nu/2-1} t_i^2 \mathrm{d}\mu(\boldsymbol{H})\mathrm{d}\boldsymbol{t}}. \tag{4}$$

It is easily seen that $\tau_{ij}(\boldsymbol{l})$ $(1 \le i, j \le p)$ is scale-invariant, that is, $\tau_{ij}(c\boldsymbol{l}) = \tau_{ij}(\boldsymbol{l})$ for any positive constant c. Furthermore τ_{ij}'s are nonnegative and

$$\sum_{j=1}^{p} \tau_{ij}(\boldsymbol{l}) = \frac{1}{\nu+2},$$

since $\sum_j h_{ij}^2 = 1$. This means that $\boldsymbol{\psi}^*(\boldsymbol{l}) = (\psi_1^*(\boldsymbol{l}), \ldots, \psi_p^*(\boldsymbol{l}))$ is an estimator which shrinks $\boldsymbol{l}/(\nu+2)$.

$\psi_i^*(\boldsymbol{l})$ $(1 \le i \le p)$ has another useful expression;

$$\psi_i^*(\boldsymbol{l}) = \left(\sum_{j=1}^{p} \tilde{\tau}_{ij}(\boldsymbol{l})\right) l_i \tag{5}$$