Journal of Multivariate Analysis 130 (2014) 27–44

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Estimations for some functions of covariance matrix in high dimension under non-normality and its applications

Tetsuto Himeno [a,*], Takayuki Yamada [b,1]

[a] *Transdisciplinary Research Integration Center, Research Organization of Information and Systems, Kamiyacho Central Place 2F, 4-3-13 Toranomon, Minato-ku, Tokyo 105-0001, Japan*

[b] *The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

ARTICLE INFO

Article history:
Received 19 July 2012
Available online 9 May 2014

AMS 2010 subject classifications:
primary 62F12
secondary 62H12

Keywords:
High dimension
Non-normal case
Covariance matrix
Estimation
Unbiasedness
Consistency

ABSTRACT

When we consider a statistical test in the high dimensional case, we often need estimators of the functions of the covariance matrix Σ. Especially, it is needed to estimate $a_2 = (1/p)\mathrm{tr}\Sigma^2$. The unbiased and consistent estimator of a_2 is proposed in preceding study when the population distribution is multivariate normal. But it is difficult to estimate in the non-normal case. So we propose the unbiased and consistent estimators for some functions of covariance matrix including a_2 under the non-normal case. Through the numerical simulation, we confirmed the accuracy of the approximation of our proposed estimators. Using proposed estimators, we proposed a test for assessing multivariate normality of the high-dimensional data.

1. Introduction

We consider the one-sample problem, that is, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be independent p-dimensional random vectors, each $\boldsymbol{x}_i$ can be expressed as

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \Sigma^{1/2}\boldsymbol{z}_i, \tag{1}$$

and $\boldsymbol{z}_i$ has a distribution F with mean vector $\mathbf{0}$ and a covariance matrix Σ, where Σ is $p \times p$ positive definite. When we deal with statistical tests of the multivariate analysis for high-dimensional data, we often need the estimator of $\mathrm{tr}\,\Sigma^2$ (e.g. Schott [11], Srivastava and Fujikoshi [13], Chen and Qin [1], Fujikoshi et al. [4], and so on). Srivastava [12] proposed unbiased and consistent estimator of $\mathrm{tr}\,\Sigma^2/p (=: a_2)$ under some conditions and the assumption that F is normal. The estimator is given by

$$\tilde{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left[\mathrm{tr}\,S^2 - \frac{1}{n}(\mathrm{tr}\,S)^2\right], \tag{2}$$

* Correspondence to: Department of Computer and Information Science, Faculty of Science and Technology, Seikei University, 3-3-1 Kichijoji-Kitamachi, Musashino-shi, Tokyo 180-8633, Japan.

E-mail address: t-himeno@st.seikei.ac.jp (T. Himeno).

1 Present address: General Studies, College of Engineering, Nihon University, 1 Nakagawara, Tokusada, Tamuramachi, Koriyama, Fukushima 963-8642, Japan.

http://dx.doi.org/10.1016/j.jmva.2014.04.020
0047-259X/

where

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i, \quad n = N-1,$$

$$S = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'.$$

Subsequently, Srivastava et al. [14] verified that this has consistency under a condition "C_S: the elements of $\boldsymbol{z} = (z_1, \ldots, z_p)' \sim F$ are independent of each other and the eighth moments of z_i $(i = 1, \ldots, p)$ are finite" even if F is not normal. Note that their estimator is not generally unbiased. Chen and Qin [1] and Chen et al. [2] also proposed other consistent estimator of a_2 under the condition "C_C : $\mathrm{E}[z_{\ell_1}^{\alpha_1} z_{\ell_2}^{\alpha_2} \cdots z_{\ell_q}^{\alpha_q}] = \mathrm{E}[z_{\ell_1}^{\alpha_1}] \cdots \mathrm{E}[z_{\ell_q}^{\alpha_q}]$ for a positive integer q such that $\sum_{i=1}^{q}\alpha_i \le 8$ and $\ell_1 \ne \ell_2 \ne \cdots \ne \ell_q$ and the fourth moments of z_i $(i = 1, \ldots, p)$ are finite". However the method of Chen and Qin has a bias and the method of Chen et al. is complicated form. Additionally, the consistency of these estimators is derived under strong moment conditions which it is difficult to satisfy without the assumption that elements of $\boldsymbol{z}$ are independent of each other. So, we propose the unbiased consistent estimators of not only $\mathrm{tr}\,\Sigma^2$ but also $(\mathrm{tr}\,\Sigma)^2$ and κ_{11} in simple form, where

$$\kappa_{ij} := \mathrm{E}[\boldsymbol{z}'\Sigma^i\boldsymbol{z}\boldsymbol{z}'\Sigma^j\boldsymbol{z}] - 2\,\mathrm{tr}\,\Sigma^{i+j} - \mathrm{tr}\,\Sigma^i\,\mathrm{tr}\,\Sigma^j. \tag{3}$$

For normal case, κ_{ij} is equal to zero (Magnus and Neudecker [8]). κ_{11} is one of the important parameters to indicate difference from the normal distribution.

The present paper is organized as follows. In Section 2, we propose these estimators and prove the unbiasedness and the consistency. In Section 3, we verify the performance of the our proposed estimators by using Monte Carlo method. In Section 4, using our proposed estimators, we provide the new test statistics testing the normality. Technical details are provided in an Appendix.

2. Unbiasedness and consistency of estimators

Firstly, we set up the asymptotic framework A1 and assumption A2:

A1 : $n/p \to c_0 \in (0, \infty)$,
A2 : $a_i := \mathrm{tr}\,\Sigma^i/p \to a_{i0} \in (0, \infty)$ (for $i = 1, \ldots, 4$).

When the estimation of $\mathrm{tr}\,\Sigma^2$ is considered, we often use $\mathrm{tr}\,S^2$ and $(\mathrm{tr}\,S)^2$. The expectations of these statistics are calculated as

$$\mathrm{E}[\mathrm{tr}\,S^2] = \frac{1}{N}\kappa_{11} + \frac{N}{N-1}\mathrm{tr}\,\Sigma^2 + \frac{1}{N-1}(\mathrm{tr}\,\Sigma)^2, \tag{4}$$

$$\mathrm{E}[(\mathrm{tr}\,S)^2] = \frac{1}{N}\kappa_{11} + \frac{2}{N-1}\mathrm{tr}\,\Sigma^2 + (\mathrm{tr}\,\Sigma)^2. \tag{5}$$

Since there are three unknown parameters in the expectations, it is impossible to estimate $\mathrm{tr}\,\Sigma^2$ by only $\mathrm{tr}\,S^2$ and $(\mathrm{tr}\,S)^2$. So we propose another statistics as

$$Q := \frac{1}{N-1}\sum_{i=1}^{N}((\boldsymbol{x}_i - \bar{\boldsymbol{x}})'(\boldsymbol{x}_i - \bar{\boldsymbol{x}}))^2. \tag{6}$$

The expectation of Q is expressed as

$$\mathrm{E}[Q] = \frac{N^2 - 3N + 3}{N^2}\kappa_{11} + \frac{2(N-1)}{N}\mathrm{tr}\,\Sigma^2 + \frac{N-1}{N}(\mathrm{tr}\,\Sigma)^2. \tag{7}$$

By solving simultaneous equations (4), (5), and (7), we obtain the following theorem.

Theorem 1. *For model (1), the unbiased estimators of* $\mathrm{tr}\,\Sigma^2$, $(\mathrm{tr}\,\Sigma)^2$, *and* κ_{11} *are obtained as*

$$\widehat{\mathrm{tr}\,\Sigma^2} = \frac{N-1}{N(N-2)(N-3)}\{(N-1)(N-2)\mathrm{tr}\,S^2 + (\mathrm{tr}\,S)^2 - NQ\},$$

$$\widehat{(\mathrm{tr}\,\Sigma)^2} = \frac{N-1}{N(N-2)(N-3)}\{2\,\mathrm{tr}\,S^2 + (N^2 - 3N + 1)(\mathrm{tr}\,S)^2 - NQ\},$$

$$\hat{\kappa}_{11} = \frac{-1}{(N-2)(N-3)}\{2(N-1)^2\,\mathrm{tr}\,S^2 + (N-1)^2(\mathrm{tr}\,S)^2 - N(N+1)Q\}.$$