Statistics and Probability Letters 129 (2017) 141–146

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A note on the unbiased estimator of Σ^2

Bu Zhou, Jia Guo *

Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore

ARTICLE INFO

Article history:
Received 5 February 2017
Received in revised form 9 May 2017
Accepted 18 May 2017
Available online 31 May 2017

MSC 2010:
62H12
62G05

Keywords:
Unbiased estimators
U-statistics
Covariance matrix
High-dimensional data

ABSTRACT

This paper gives simple and intuitive derivations of three equivalent forms of a distribution-free and unbiased estimator of the squared covariance matrix $\boldsymbol{\Sigma}^2$. Particularly, computationally efficient forms of the unbiased estimators of $\boldsymbol{\Sigma}^2$ and its trace are derived from the computationally intensive U-statistic forms.

1. Introduction

Suppose we have n independently distributed p-variate samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ with a common mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, we are interested in the estimation of the squared covariance matrix $\boldsymbol{\Sigma}^2$. This work is inspired by the estimation of the trace of the squared covariance matrix $\mathrm{tr}(\boldsymbol{\Sigma}^2)$, which is used in many high-dimensional inference problems, such as the high-dimensional mean and covariance testing problems, see for example, Bai and Saranadasa (1996), Chen and Qin (2010), and Srivastava and Yanagihara (2010) among others. Let $\mathbf{S} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top/(n-1)$ be the sample covariance matrix, in the classical setting when the dimension p is fixed, a very natural plug-in estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ is $\mathrm{tr}(\mathbf{S}^2)$. However, $\mathrm{tr}(\mathbf{S}^2)$ is biased and is not consistent when p diverges because the bias of $\mathrm{tr}(\mathbf{S}^2)$ depends on p. When the samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are i.i.d. (independently and identically distributed) from a multivariate normal distribution, an unbiased estimator can be easily found using the moments formulas of Wishart matrix. An unbiased and translation-invariant estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ for normally distributed samples is given, for example, in Bai and Saranadasa (1996) (see also (3.18) of Zhang and Xu, 2009) and is used in the estimation of the variance of their two-sample test statistic. This unbiased estimator is in fact a UMVUE (uniformly minimum variance unbiased estimator) of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ under the normality assumption because it is a function of $\mathbf{S}$ only (see also Lemma 2 of Hu et al., 2017). However, the unbiased estimator is generally biased for non-normal data and further assumptions are needed for its asymptotic unbiasedness. Some other asymptotically unbiased estimators of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ are proposed, for example, in Ahmad et al. (2008) and Chen and Qin (2010), but these estimators also require additional conditions for asymptotic unbiasedness and are usually not translation-invariant.

In high-dimensional analysis, the unbiasedness of an estimator is a very desirable property because any small bias depending on p may no longer be negligible when p diverges. Besides, we want the estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ to be translation-invariant in the sense that it does not depend on the mean vector $\boldsymbol{\mu}$ because in many cases when we need the estimation of

* Corresponding author.
E-mail addresses: bu.zhou@u.nus.edu (B. Zhou), jia.guo@u.nus.edu (J. Guo).

http://dx.doi.org/10.1016/j.spl.2017.05.014
0167-7152/

$\mathrm{tr}(\boldsymbol{\Sigma}^2)$, $\boldsymbol{\mu}$ is also unknown. Another important factor to be considered is the computational efficiency, as when both n and p are large, the computation times for inefficient estimators can be intolerable.

In this paper, we derive three different forms of the unbiased estimator of $\boldsymbol{\Sigma}^2$ under only the independence and the common mean and covariance assumptions. In particular, a computationally efficient form of the unbiased estimator of $\boldsymbol{\Sigma}^2$ is derived from the computationally intensive U-statistic form. The derivations are simple, intuitive and extendable, only relying on some basic ideas of U-statistics and elementary calculations. Based on the three different forms of the unbiased estimator of $\boldsymbol{\Sigma}^2$, we can obtain three forms, including a computationally efficient form, of the unbiased and translation-invariant estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$, that appear in Hu et al. (2017), Chen et al. (2010) and Yamada and Himeno (2015), respectively, and establish their equivalence automatically. Especially, this paper provides a new and intuitive way to derive the computationally efficient form of the unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ given by Yamada and Himeno (2015), and thus unifies the three unbiased and translation-invariant estimators of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ proposed respectively by Chen et al. (2010), Hu et al. (2017) and Yamada and Himeno (2015), from a U-statistics perspective.

2. Main results

An unbiased estimator can be directly given in the form of a U-statistic. Note that $\boldsymbol{\Sigma}^2 = \mathrm{E}\{(\mathbf{x}_1-\mathbf{x}_3)(\mathbf{x}_1-\mathbf{x}_4)^\top(\mathbf{x}_2-\mathbf{x}_5)(\mathbf{x}_2-\mathbf{x}_6)^\top\}$, by averaging all possible combinations, the U-statistic for estimating $\boldsymbol{\Sigma}^2$ is

$$\widehat{\boldsymbol{\Sigma}^2}_{(1)} = \frac{1}{P_6}\sum_{s,t,i,j,k,l}^{*}\left\{(\mathbf{x}_s-\mathbf{x}_i)(\mathbf{x}_s-\mathbf{x}_j)^\top(\mathbf{x}_t-\mathbf{x}_k)(\mathbf{x}_t-\mathbf{x}_l)^\top\right\}, \tag{1}$$

where the "hollow sum" $\sum^{*}_{i_1,\ldots,i_m}$ denotes the summation over mutually distinct indices $i_1,\ldots,i_m \in \{1,\ldots,n\}$, and $P_m = n\times\cdots\times(n-m+1)$. By linearity of the trace and the expectation operators, we obtain the first form of the unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$,

$$\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(1)} = \mathrm{tr}(\widehat{\boldsymbol{\Sigma}^2}_{(1)}) = \frac{1}{P_6}\sum_{s,t,i,j,k,l}^{*}\left\{(\mathbf{x}_t-\mathbf{x}_l)^\top(\mathbf{x}_s-\mathbf{x}_i)(\mathbf{x}_s-\mathbf{x}_j)^\top(\mathbf{x}_t-\mathbf{x}_k)\right\}, \tag{2}$$

which is a U-statistic for estimating $\mathrm{tr}(\boldsymbol{\Sigma}^2)$. The form $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(1)}$ was used by Hu et al. (2017) for their MANOVA test. The unbiasedness of $\widehat{\boldsymbol{\Sigma}^2}_{(1)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(1)}$ does not depend on the underlying distribution of the samples. What is more, $\widehat{\boldsymbol{\Sigma}^2}_{(1)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(1)}$ are translation-invariant because the factors like $\mathbf{x}_s-\mathbf{x}_i$ in (1) and (2) are all translation-invariant. Obviously, $\widehat{\boldsymbol{\Sigma}^2}_{(1)}$ is computationally intensive because the summation $\sum^{*}_{s,t,i,j,k,l}$ is of order $O(n^6)$. But notice that for any term in the expansion of $(\mathbf{x}_s-\mathbf{x}_i)(\mathbf{x}_s-\mathbf{x}_j)^\top(\mathbf{x}_t-\mathbf{x}_k)(\mathbf{x}_t-\mathbf{x}_l)^\top$, we have four unique indices at most, for example, $\mathbf{x}_i\mathbf{x}_j^\top\mathbf{x}_k\mathbf{x}_l^\top$, so the order of the summation can be reduced to $O(n^4)$ by expanding the kernel $(\mathbf{x}_s-\mathbf{x}_i)(\mathbf{x}_s-\mathbf{x}_j)^\top(\mathbf{x}_t-\mathbf{x}_k)(\mathbf{x}_t-\mathbf{x}_l)^\top$. This is equivalent to an alternative approach of constructing the unbiased estimator of $\boldsymbol{\Sigma}^2$ based on a linear combination of U-statistics. Note $\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{x}_1\mathbf{x}_1^\top)-\boldsymbol{\mu}\boldsymbol{\mu}^\top$, we have $\boldsymbol{\Sigma}^2 = \left\{\mathrm{E}(\mathbf{x}_1\mathbf{x}_1^\top)\right\}^2 + \left(\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)^2 - \mathrm{E}(\mathbf{x}_1\mathbf{x}_1^\top)\boldsymbol{\mu}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top\mathrm{E}(\mathbf{x}_1\mathbf{x}_1^\top)$ by expanding $\boldsymbol{\Sigma}^2$. Replacing the four terms in the expansion of $\boldsymbol{\Sigma}^2$ by their corresponding U-statistics, we have the following estimator of $\boldsymbol{\Sigma}^2$,

$$\widehat{\boldsymbol{\Sigma}^2}_{(2)} = \frac{1}{P_2}\sum_{i,j}^{*}\left(\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_j\mathbf{x}_j^\top\right) + \frac{1}{P_4}\sum_{i,j,k,l}^{*}\left(\mathbf{x}_i\mathbf{x}_j^\top\mathbf{x}_k\mathbf{x}_l^\top\right) - \frac{1}{P_3}\sum_{i,j,k}^{*}\left(\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_j\mathbf{x}_k^\top\right) - \frac{1}{P_3}\sum_{i,j,k}^{*}\left(\mathbf{x}_i\mathbf{x}_j^\top\mathbf{x}_k\mathbf{x}_k^\top\right), \tag{3}$$

and the corresponding estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$,

$$\begin{aligned}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(2)} &= \mathrm{tr}\left\{\widehat{\boldsymbol{\Sigma}^2}_{(2)}\right\}\\ &= \frac{1}{P_2}\sum_{i,j}^{*}\left(\mathbf{x}_j^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_j\right) + \frac{1}{P_4}\sum_{i,j,k,l}^{*}\left(\mathbf{x}_l^\top\mathbf{x}_i\mathbf{x}_j^\top\mathbf{x}_k\right) - \frac{2}{P_3}\sum_{i,j,k}^{*}\left(\mathbf{x}_k^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_j\right).\end{aligned} \tag{4}$$

The above estimator $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(2)}$ was used in Chen et al. (2010) for constructing their test statistics. The equivalence of $\widehat{\boldsymbol{\Sigma}^2}_{(1)}$ and $\widehat{\boldsymbol{\Sigma}^2}_{(2)}$ can be easily shown by expanding $(\mathbf{x}_s-\mathbf{x}_i)(\mathbf{x}_s-\mathbf{x}_j)^\top(\mathbf{x}_t-\mathbf{x}_k)(\mathbf{x}_t-\mathbf{x}_l)^\top$ (see a complete proof in Appendix A), so we have the following fact.

Fact 1. *$\widehat{\boldsymbol{\Sigma}^2}_{(1)} = \widehat{\boldsymbol{\Sigma}^2}_{(2)}$, and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(1)} = \widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(2)}$.*

By Fact 1, we immediately know that $\widehat{\boldsymbol{\Sigma}^2}_{(2)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}_{(2)}$ are also translation-invariant.

The second form $\widehat{\boldsymbol{\Sigma}^2}_{(2)}$ given in (3) still needs a summation of order $O(n^4)$, which is also very slow to compute when n is large. A common way to reduce the computation of hollow sums in U-statistics is to represent them in terms of complete sums without any indices restrictions as in Lemma 1 in Appendix A. One may directly simplify the hollow sums in (3) using this idea. However, the computation can be significantly reduced by exploiting the translation-invariant property of the