doi：10.3969/j.issn.0490-6756.2020.04.003

协方差矩阵的乘积及其迹的估计

沈洁琼[1,2]，周　杰[1]，陈　晨[1]

(1. 四川大学数学学院，成都 610064；2. 浙大宁波理工学院计算机与数据工程学院，宁波 315100)

摘　要：对于两个不同总体的协方差矩阵 Σ_1 和 Σ_2，估计其乘积 $\Sigma_1\Sigma_2$ 及乘积的迹 $\mathrm{tr}(\Sigma_1\Sigma_2)$ 是统计推断问题的关键步骤. 本文首先构造了 $\Sigma_1\Sigma_2$ 的几个等价估计，并对任意的正整数 m，n 建立了 $\Sigma_1^m\Sigma_2^n$ 和 $(\Sigma_1\Sigma_2)^m$ 的无偏估计. 其次，利用 $\Sigma_1\Sigma_2$ 的等价估计，本文证明 $\mathrm{tr}(\Sigma_1\Sigma_2)$ 的多个常用估计量是相等的. 本文最后证明了两个常用的检验统计量（被用于检验两个协方差矩阵是否相等）是渐近等价的.

关键词：协方差矩阵；估计；统计推断；等价

中图分类号：O212.4　　**文献标识码**：A　　**文章编号**：0490-6756(2020)04-0635-07

On the estimation for product of covariance matrices and its trace

SHEN Jie-Qiong[1,2], *ZHOU Jie*[1], *CHEN Chen*[1]

(1. School of Mathematics, Sichuan University, Chengdu 610064, China;
2. School of Computer and Data Engineering, Ningbo Tech University, Ningbo 315100, China)

Abstract: In many statistical inference problems involving two populations respectively with covariance matrices Σ_1 and Σ_2, the product $\Sigma_1\Sigma_2$ and its trace $\mathrm{tr}(\Sigma_1\Sigma_2)$ need to be estimated. Firstly, we construct some equivalent estimators of $\Sigma_1\Sigma_2$ and unbiased estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ for any positive integers m and n. Secondly, by using the equivalent estimators of $\Sigma_1\Sigma_2$, we show that some existing estimators of $\mathrm{tr}(\Sigma_1\Sigma_2)$ are equal. Finally, it is proved that two existing test statistics for testing the equality of two covariance matrices are asymptotically equivalent.

Keywords: Covariance matrix; Estimation; Statistical inference; Equivalence

1　Introduction

Consider two p-dimensional populations respectively with the mean vector μ_i and covariance matrix Σ_i, $i=1,2$. The estimate of $\mathrm{tr}(\Sigma_1\Sigma_2)$ is essential and frequently encountered in multivariate statistical analysis, in particular, in two-sample covariance matrices testing problem. For example, one wants to test the hypotheses

$$H_0:\Sigma_1=\Sigma_2\ vs.\ H_1:\Sigma_1\neq\Sigma_2 \qquad (1)$$

In the literature, the distance function between the null and alternative hypotheses is usually given by

$$d:=\mathrm{tr}(\Sigma_1-\Sigma_2)^2=\mathrm{tr}\Sigma_1^2+\mathrm{tr}\Sigma_2^2-2\mathrm{tr}(\Sigma_1\Sigma_2),$$

which indicates that $d=0$ if and only if the hypothesis H_0 is true. Therefore, one can construct the test statistic for the testing problem (1) by employing the estimators of $\mathrm{tr}\Sigma_1^2$, $\mathrm{tr}\Sigma_2^2$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$. The estimators of $\mathrm{tr}\Sigma_1^2$ have been exten-

收稿日期：2019-01-13
基金项目：国家自然科学基金(61374027，11871357)；四川省应用基础研究项目(2019YJ0122)
作者简介：沈洁琼(1985－)，女，浙江宁波人，博士，主要研究领域为数理统计. E-mail：jieqiongshen123@163.com
通讯作者：周杰. E-mail：jzhou@scu.edu.cn

sively studied. For more details, we refer to Refs. [1-4]. Therefore, we aim to seek some suitable estimators of $\Sigma_1\Sigma_2$ and then directly obtain the corresponding estimators of $\mathrm{tr}(\Sigma_1\Sigma_2)$.

Let $\{X_{ij}\}_{j=1}^{N_i}$ be independent and identically distributed sample from the i-th population with sample sizes N_i, $i=1, 2$, and $X_{1j}'s$ and $X_{2j}'s$ are independent. The intuitive and natural estimators of $\Sigma_1\Sigma_2$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$ are respectively given by

$$A_1 \coloneqq S_1 S_2,\ B_1 \coloneqq \mathrm{tr}(S_1 S_2) \tag{2}$$

where

$$S_i = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_i)^T,$$

$$\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}X_{ij},\ i=1,2.$$

Note that A_1 and B_1 are unbiased and location-invariant (i. e., not depending on the mean vectors) estimators of $\Sigma_1\Sigma_2$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$, respectively, and the latter has been employed to construct the test statistic for the testing problem (1) in Ref. [3]. We also remark here that computing A_1 and B_1 require $N_1p^2+N_2p^2+p^3$ and $N_1p^2+N_2p^2+p^2$ multiplications, respectively.

It is worth pointing out that, for the normal populations, $(N_i-1)S_i$ follows the p-dimensional Wishart distribution with the degree of freedom N_i-1 and covariance matrix Σ_i, and then we can obtain some essential properties of S_1S_2 by utilizing the properties of Wishart distribution. However, such results do not hold when the populations dispense with the normality assumption. Besides, $S_1^mS_2^n$ and $(S_1S_2)^m$ may be biased estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ for some positive integers m and n, respectively.

The main results are provided in Section 2: i) some equivalent estimators of $\Sigma_1\Sigma_2$ are established by the basic idea of U-statistics and three commonly used estimators of $\mathrm{tr}(\Sigma_1\Sigma_2)$ respectively given in Refs. [5], [3] and [6] are found to be exactly the same but with different computational complexities; ii) the unbiased and location-invariant estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ are constructed for any positive integers m and n; iii) the test statistics proposed in Ref. [5] and Ref. [3] for the testing problem (1) are shown to be asymptotically equivalent. Section 3 gives a conclusion.

2 Equivalent estimators of $\Sigma_1\Sigma_2$ and tr $(\Sigma_1\Sigma_2)$

A straightforward calculation shows that

$$\Sigma_1 = E((X_{11}-X_{12})(X_{11}-X_{13})^{\mathrm{T}})$$

and

$$\Sigma_2 = E((X_{21}-X_{22})(X_{21}-X_{23})^{\mathrm{T}}),$$

which leads to

$$\Sigma_1\Sigma_2 = E((X_{11}-X_{12})(X_{11}-X_{13})^T (X_{21}-X_{22})(X_{21}-X_{23})^T).$$

By covering all possible combinations, which is the essence of the U-statistics formulation, the second estimator of $\Sigma_1\Sigma_2$ is given as

$$A_2 \coloneqq \frac{1}{P_{N_1}^3 P_{N_2}^3}\sum_{i,j,k}^{*}\sum_{l,s,t}^{*}(X_{1i}-X_{1j})(X_{1i}-X_{1k})^T(X_{2l}-X_{2s})(X_{2l}-X_{2t})^T \tag{3}$$

Hereinafter, $\sum^{*}$ denotes the summation over all mutually different indices and $P_a^b = a(a-1)\cdots(a-b+1)$. According to A_2 given by (3), we can obtain an estimator of $\mathrm{tr}(\Sigma_1\Sigma_2)$ by utilizing the properties of the trace:

$$B_2 \coloneqq \mathrm{tr}A_2 = \frac{1}{P_{N_1}^3 P_{N_2}^3}\sum_{i,j,k}^{*}\sum_{l,s,t}^{*}(X_{1i}-X_{1k})^{\mathrm{T}}(X_{2l}-X_{2s})(X_{2l}-X_{2t})^{\mathrm{T}}(X_{1i}-X_{1j}).$$

Note that A_2 and B_2 are unbiased and location-invariant estimators of $\Sigma_1\Sigma_2$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$, respectively. Obviously, computing A_2 and B_2 need $N_1(N_1-1)(N_1-2)p^2+N_2(N_2-1)(N_2-2)p^2+p^3$ and $N_1(N_1-1)(N_1-2)p^2+N_2(N_2-1)(N_2-2)p^2+p^2$ multiplications, respectively.

Noticing that

$$E((X_{11}-X_{12})(X_{11}-X_{12})^T)=2\Sigma_1$$

and

$$E((X_{21}-X_{22})(X_{21}-X_{22})^T)=2\Sigma_2,$$

we also have

$$\Sigma_1\Sigma_2 = \frac{1}{4}E((X_{11}-X_{12})(X_{11}-X_{12})^T (X_{21}-X_{22})(X_{21}-X_{22})^T).$$

Hence, the third unbiased and location-invariant estimator of $\Sigma_1\Sigma_2$ is proposed as

$$A_3 \coloneqq \frac{1}{4P_{N_1}^2 P_{N_2}^2} \sum_{i,j}^{*} \sum_{k,l}^{*} (X_{1i} - X_{1j}) (X_{1i} - X_{1j})^T (X_{2k} - X_{2l}) (X_{2k} - X_{2l})^T \quad (4)$$

Compatibly with A_3 given by (4), the third estimator of tr($\Sigma_1\Sigma_2$) can be repressed as

$$B_3 \coloneqq \mathrm{tr}A_3 = \frac{1}{4P_{N_1}^2 P_{N_2}^2} \sum_{i,j}^{*} \sum_{k,l}^{*} (X_{1i} - X_{1j})^T (X_{2k} - X_{2l}) (X_{2k} - X_{2l})^T (X_{1i} - X_{1j}).$$

In Ref. [6], by virtue of B_3, the test statistics for testing multi-sample sphericity and identity of covariance matrices are constructed. Note that computing A_3 and B_3 take $N_1(N_1-1)p^2+N_2(N_2-1)p^2+p^3$ and $N_1(N_1-1)p^2+N_2(N_2-1)p^2+p^2$ multiplications, respectively.

From

$$\Sigma_1 = E(X_{11}X_{11}^T) - \mu_1\mu_1^T$$

and

$$\Sigma_2 = E(X_{21}X_{21}^T) - \mu_2\mu_2^T,$$

we have

$$\begin{aligned}\Sigma_1\Sigma_2 = {} & (E(X_{11}X_{11}^T) - \mu_1\mu_1^T)(E(X_{21}X_{21}^T) - \mu_2\mu_2^T) = \\ & E(X_{11}X_{11}^T)E(X_{21}X_{21}^T) - E(X_{11}X_{11}^T)\mu_2\mu_2^T - \\ & \mu_1\mu_1^T E(X_{21}X_{21}^T) + \mu_1\mu_1^T\mu_2\mu_2^T \end{aligned} \quad (5)$$

By replacing four terms on the right hand side of (5) with their corresponding U-statistics respectively, the fourth unbiased and location-invariant estimator of $\Sigma_1\Sigma_2$ can be obtained as

$$\begin{aligned} A_4 \coloneqq {} & \frac{1}{P_{N_1}^1 P_{N_2}^1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} X_{1i}X_{1i}^T X_{2j}X_{2j}^T - \\ & \frac{1}{P_{N_1}^1 P_{N_2}^2} \sum_{i=1}^{N_1} \sum_{j,k}^{*} X_{1i}X_{1i}^T X_{2j}X_{2k}^T - \\ & \frac{1}{P_{N_1}^2 P_{N_2}^1} \sum_{i,j}^{*} \sum_{k=1}^{N_2} X_{1i}X_{1j}^T X_{2k}X_{2k}^T + \\ & \frac{1}{P_{N_1}^2 P_{N_2}^2} \sum_{i,j}^{*} \sum_{k,l}^{*} X_{1i}X_{1j}^T X_{2k}X_{2l}^T \end{aligned} \quad (6)$$

Correspondingly, the compatible estimator of tr ($\Sigma_1\Sigma_2$) is given as

$$\begin{aligned} B_4 \coloneqq \mathrm{tr}A_4 = {} & \frac{1}{P_{N_1}^1 P_{N_2}^1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} X_{1i}^T X_{2j}X_{2j}^T X_{1i} - \\ & \frac{1}{P_{N_1}^1 P_{N_2}^2} \sum_{i=1}^{N_1} \sum_{j,k}^{*} X_{1i}^T X_{2j}X_{2k}^T X_{1i} - \\ & \frac{1}{P_{N_1}^2 P_{N_2}^1} \sum_{i,j}^{*} \sum_{k=1}^{N_2} X_{1j}^T X_{2k}X_{2k}^T X_{1i} + \\ & \frac{1}{P_{N_1}^2 P_{N_2}^2} \sum_{i,j}^{*} \sum_{k,l}^{*} X_{1j}^T X_{2k}X_{2l}^T X_{1i} \end{aligned} \quad (7)$$

The estimator B_4 is proposed in Ref. [5] to construct the test statistic for the testing problem (1). We mention that computing A_4 and B_4, which are the linear combination of U-statistics, need $2N_1^2p^2+2N_2^2p^2+4p^3$ and $2N_1^2p^2+2N_2^2p^2+4p^2$ multiplications, respectively. For easy to compare, we summarize the computational complexities of different forms of the estimators in Table 1.

Tab. 1　Computational complexity of the estimators

Estimator	Multiplication
A_1	$N_1p^2+N_2p^2+p^3$
A_2	$N_1(N_1-1)(N_1-2)p^2+N_2(N_2-1)(N_2-2)p^2+p^3$
A_3	$N_1(N_1-1)p^2+N_2(N_2-1)p^2+p^3$
A_4	$2N_1^2p^2+2N_2^2p^2+4p^3$
B_1	$N_1p^2+N_2p^2+p^2$
B_2	$N_1(N_1-1)(N_1-2)p^2+N_2(N_2-1)(N_2-2)p^2+p^2$
B_3	$N_1(N_1-1)p^2+N_2(N_2-1)p^2+p^2$
B_4	$2N_1^2p^2+2N_2^2p^2+4p^2$

Theorem 2.1　$A_1=A_2=A_3=A_4$ and then $B_1=B_2=B_3=B_4$.

Proof　It is sufficient to prove $A_1=A_2=A_3=A_4$. Let $\tilde{X}_{ij}=X_{ij}-\bar{X}_i$, $i=1,2$, $j=1,\cdots N_i$. Replacing X_{ij} in (6) with $\tilde{X}_{ij}$ and using the property of location-invariance of A_4, we can rewrite A_4 as

$$\begin{aligned} A_4 \coloneqq {} & \frac{1}{P_{N_1}^1 P_{N_2}^1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{X}_{1i}\tilde{X}_{1i}^T \tilde{X}_{2j}\tilde{X}_{2j}^T - \\ & \frac{1}{P_{N_1}^1 P_{N_2}^2} \sum_{i=1}^{N_1} \sum_{j,k}^{*} \tilde{X}_{1i}\tilde{X}_{1i}^T \tilde{X}_{2j}\tilde{X}_{2k}^T - \\ & \frac{1}{P_{N_1}^2 P_{N_2}^1} \sum_{i,j}^{*} \sum_{k=1}^{N_2} \tilde{X}_{1i}\tilde{X}_{1j}^T \tilde{X}_{2k}\tilde{X}_{2k}^T + \\ & \frac{1}{P_{N_1}^2 P_{N_2}^2} \sum_{i,j}^{*} \sum_{k,l}^{*} \tilde{X}_{1i}\tilde{X}_{1j}^T \tilde{X}_{2k}\tilde{X}_{2l}^T \end{aligned} \quad (8)$$

On the other hand, from $\sum_{j=1}^{N_i} \tilde{X}_{ij} = 0$, $i=1,2$, we have

$$\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{X}_{1i}\tilde{X}_{1i}^T \tilde{X}_{2j}\tilde{X}_{2j}^T = (N_1-1)(N_2-1)S_1S_2 \quad (9)$$

$$\sum_{i=1}^{N_1}\sum_{j,k}^{*}\widetilde{X}_{1i}\widetilde{X}_{1i}^T\widetilde{X}_{2j}\widetilde{X}_{2k}^T = \sum_{i=1}^{N_1}\widetilde{X}_{1i}\widetilde{X}_{1i}^T(\sum_{j,k=1}^{N_2}\widetilde{X}_{2j}\widetilde{X}_{2k}^T - \sum_{j=1}^{N_2}\widetilde{X}_{2j}\widetilde{X}_{2j}^T) = -(N_1-1)(N_2-1)S_1S_2 \tag{10}$$

$$\sum_{i,j}^{*}\sum_{k=1}^{N_2}\widetilde{X}_{1i}\widetilde{X}_{1j}^T\widetilde{X}_{2k}\widetilde{X}_{2k}^T = (\sum_{i,j=1}^{N_1}\widetilde{X}_{1i}\widetilde{X}_{1j}^T - \sum_{i=1}^{N_1}\widetilde{X}_{1i}\widetilde{X}_{1i}^T)\sum_{k=1}^{N_2}\widetilde{X}_{2k}\widetilde{X}_{2k}^T = -(N_1-1)(N_2-1)S_1S_2 \tag{11}$$

$$\sum_{i,j}^{*}\sum_{k,l}^{*}\widetilde{X}_{1i}\widetilde{X}_{1j}^T\widetilde{X}_{2k}\widetilde{X}_{2l}^T = (\sum_{i,j=1}^{N_1}\widetilde{X}_{1i}\widetilde{X}_{1j}^T - \sum_{i=1}^{N_1}\widetilde{X}_{1i}\widetilde{X}_{1i}^T)(\sum_{k,l=1}^{N_2}\widetilde{X}_{2k}\widetilde{X}_{2l}^T - \sum_{k=1}^{N_2}\widetilde{X}_{2k}\widetilde{X}_{2k}^T) = (N_1-1)(N_2-1)S_1S_2 \tag{12}$$

Substituting (9),(10) into (8) yields $A_4 = A_1$.

Some straightforward calculations indicate that $A_2 = A_4$ and $A_3 = A_4$, respectively. This completes the proof of this theorem.

***Remark* 1** We recall that the matrix estimators A_2, A_3 and A_4 are constructed by the idea of U-statistics. It particularly indicates that some desirable properties such as consistency of A_i, $i=2, 3, 4$, can be obtained conveniently based on the theory of U-statistics which does not depend on the underlying distributions.

***Remark* 2** Note that the estimators B_1, B_3 and B_4 have been directly utilized by Refs. [3], [6] and [5] respectively. However, to our best knowledge, there do not exist literatures presenting the relationship among these estimators. Theorem 2.1 asserts that these estimators, though having different forms and computational complexities, are exactly the same as $\mathrm{tr}(S_1S_2)$.

3 Unbiased estimators for more general case

Usually, the test statistic for testing problem (1) is based on an unbiased estimator of some given function $\widetilde{d}$ that is employed to distinguish the null and alternative hypotheses. We would like to mention that there are many candidates for $\widetilde{d}$ to distinguish H_0 and H_1, for instance,

$$\widetilde{d}_1 = \mathrm{tr}(\Sigma_1-\Sigma_2)^2, \widetilde{d}_2 = \mathrm{tr}(\Sigma_1-\Sigma_2)^4,$$

$$\widetilde{d}_3 = \mathrm{tr}(\Sigma_1\log\Sigma_1 - \Sigma_1\log\Sigma_2 - \Sigma_1 + \Sigma_2)$$

$$\text{or } \widetilde{d}_4 = \log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) - p,$$

where, $\widetilde{d}_1$ is employed in Ref. [5] and Ref. [3], $\widetilde{d}_2$ is motivated by Ref. [8], and $\widetilde{d}_3$ and $\widetilde{d}_4$ are Von Neumann divergence and LogDet divergence respectively, which are given in Ref. [9]. Some straightforward calculations show that, whenever we design the test statistic based on $\widetilde{d}_2$, $\widetilde{d}_3$ or $\widetilde{d}_4$, we need to estimate $\mathrm{tr}(\Sigma_1\Sigma_2^2)$, $\mathrm{tr}((\Sigma_1\Sigma_2)^2)$, etc. This motivates us to establish the unbiased estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ for both theoretical and practical purposes.

Note that, for any given positive integers m and n, $S_1^mS_2^n$ and $(S_1S_2)^m$ may be biased estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ respectively. Fortunately, we can obtain the unbiased estimators of $\Sigma_1^m\Sigma_2^n$ and $(\Sigma_1\Sigma_2)^m$ along the lines of the constructed methods of A_i, $i=2, 3, 4$.

For mutually different indices $i_1, \cdots, i_m, j_1, \cdots, j_m$ and different indices $k_1, \cdots, k_n, l_1, \cdots, l_n$, we have

$$E(\prod_{a=1}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T) = (2\Sigma_1)^m$$

and

$$E(\prod_{b=1}^{n}(X_{2k_b}-X_{2l_b})(X_{2k_b}-X_{2l_b})^T) = (2\Sigma_2)^n.$$

Therefore,

$$\Sigma_1^m\Sigma_2^n = \frac{1}{2^{m+n}}E(\prod_{a=1}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T\prod_{b=1}^{n}(X_{2k_b}-X_{2l_b})(X_{2k_b}-X_{2l_b})^T).$$

An unbiased and location-invariant estimator of $\Sigma_1^m\Sigma_2^n$ can be constructed as

$$\widehat{\Sigma_1^m\Sigma_2^n} = \frac{1}{2^{m+n}P_{N_1}^{2m}P_{N_2}^{2n}}\cdot \sum^{*}\sum^{*}(\prod_{a=1}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T\cdot \prod_{b=1}^{n}(X_{2k_b}-X_{2l_b})(X_{2k_b}-X_{2l_b})^T) \tag{13}$$

Compatibly with $\widehat{\Sigma_1^m\Sigma_2^n}$, the estimator of $\mathrm{tr}\,(\Sigma_1^m\Sigma_2^n)$ is

$$\mathrm{tr}(\widehat{\Sigma_1^m\Sigma_2^n}) = \frac{1}{2^{m+n}P_{N_1}^{2m}P_{N_2}^{2n}}\sum^{*}\sum^{*} (X_{1i_1}-X_{1j_1})^T\Big(\prod_{a=2}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T\Big)\Big(\prod_{b=1}^{n}(X_{2k_b}-X_{2l_b})(X_{2k_b}-X_{2l_b})^T\Big)(X_{1i_1}-X_{1j_1}) \quad (14)$$

Similarly, we can obtain the unbiased and location-invariant estimators of $(\Sigma_1\Sigma_2)^m$, $\mathrm{tr}\,((\Sigma_1\Sigma_2)^m)$ and $(\mathrm{tr}\,(\Sigma_1\Sigma_2))^m$ as follows.

$$\widehat{(\Sigma_1\Sigma_2)^m} = \frac{1}{4^mP_{N_1}^{2m}P_{N_2}^{2m}}\sum^{*}\sum^{*}\Big(\prod_{a=1}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T(X_{2k_a}-X_{2l_a})(X_{2k_a}-X_{2l_a})^T\Big),$$

$$\mathrm{tr}(\widehat{(\Sigma_1\Sigma_2)^m}) = \frac{1}{4^mP_{N_1}^{2m}P_{N_2}^{2m}}\sum^{*}\sum^{*}(X_{1i_1}-X_{1j_1})^T(X_{2k_1}-X_{2l_1})(X_{2k_1}-X_{2l_1})^T\cdot \Big(\prod_{a=2}^{m}(X_{1i_a}-X_{1j_a})(X_{1i_a}-X_{1j_a})^T(X_{2k_a}-X_{2l_a})(X_{2k_a}-X_{2l_a})^T\Big)(X_{1i_1}-X_{1j_1}),$$

$$\widehat{(\mathrm{tr}(\Sigma_1\Sigma_2))^m} = \frac{1}{4^mP_{N_1}^{2m}P_{N_2}^{2m}}\sum^{*}\sum^{*}\Big(\prod_{a=1}^{m}(X_{1i_a}-X_{1j_a})^T(X_{2k_a}-X_{2l_a})(X_{2k_a}-X_{2l_a})^T(X_{1i_a}-X_{1j_a})\Big).$$

***Remark* 3** When $m=1$ and $n=0$, it is clear that $\widehat{\Sigma_1^m\Sigma_2^n}$ given in (13) reduces to the sample covariance S_1, and $\mathrm{tr}(\widehat{\Sigma_1^m\Sigma_2^n})$ given in (14) arrives at $\mathrm{tr}S_1$ as well as the unbiased estimator of $\mathrm{tr}\,\Sigma_1$ proposed by Ref. [2] (see equation (2.3) in Ref. [2]).

4 Relationship between two existing test statistics

For the testing Problem (1), Ref. [5] constructed the following test statistic:

$$T_{LC} = \frac{\mathrm{tr}\widehat{\Sigma_1^2}_{(LC)} + \mathrm{tr}\widehat{\Sigma_2^2}_{(LC)} - 2B_4}{\frac{2}{N_2}\mathrm{tr}\widehat{\Sigma_1^2}_{(LC)} + \frac{2}{N_1}\mathrm{tr}\widehat{\Sigma_2^2}_{(LC)}},$$

where B_4 is given by (7) and

$$\mathrm{tr}\,\widehat{\Sigma_i^2}_{(LC)} = \frac{1}{P_{N_i}^2}\sum_{k,r}^{*}(X_{ik}^TX_{ir})^2 - \frac{2}{P_{N_i}^3}\sum_{k,r,l}^{*}X_{ik}^TX_{ir}X_{ir}^TX_{il} + \frac{1}{P_{N_i}^4}\sum_{k,r,l,s}^{*}X_{ik}^TX_{ir}X_{il}^TX_{is},\ i=1,2.$$

Moreover, Ref. [5] showed that $T_{LC}\xrightarrow{D}N(0,1)$ under H_0 as $\min\{N_1,N_2,p\}\to\infty$.

Later, for same testing problem, another test statistic was proposed in Ref. [3] as

$$T_S = \frac{\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + \mathrm{tr}\widehat{\Sigma_2^2}_{(S)} - 2B_1}{\frac{2}{n_2}\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + \frac{2}{n_1}\mathrm{tr}\widehat{\Sigma_2^2}_{(S)}},$$

where $n_i=N_i-1$, B_1 is given by (2) and

$$\mathrm{tr}\widehat{\Sigma_i^2}_{(S)} = \frac{1}{f}((N_i-2)(N_i-1)^3\cdot \mathrm{tr}S_i^2 - N_i(N_i-1)\mathrm{tr}D_i^2 + (N_i-1)^2(\mathrm{tr}S_i)^2)$$

with $f=N_i(N_i-1)(N_i-2)(N_i-3)$ and

$$D_i=\mathrm{diag}(\tilde{X}_{i1}^T\tilde{X}_{i1},\cdots,\tilde{X}_{iN_i}^T\tilde{X}_{iN_i}),\ i=1,2.$$

Under the assumption

$$N_i=O(p^{\varepsilon}),\ \frac{1}{2}<\varepsilon<1,\ i=1,2 \quad (15)$$

Ref. [3] proved that $T_S\xrightarrow{D}N(0,1)$ under H_0 as $\min\{N_1,N_2,p\}$ tends to infinity.

Next, we show that the test statistics T_{LC} and T_S are asymptotically equivalent.

Theorem 4.1 $T_S<T_{LC}$ for a fixed pair (N_1, N_2), and $T_S-T_{LC}\to 0$ as $\min\{N_1,N_2\}\to\infty$.

Proof From the fact that $\mathrm{tr}\widehat{\Sigma_i^2}_{(LC)} = \mathrm{tr}\widehat{\Sigma_i^2}_{(S)}$ for $i=1,2$ given in Ref. [7] and $B_4=B_1$ given in Theorem 2.1, a straightforward derivation shows that $T_S=\beta T_{LC}$, where

$$\beta = \frac{\frac{2}{N_2}\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + \frac{2}{N_1}\mathrm{tr}\widehat{\Sigma_2^2}_{(S)}}{\frac{2}{n_2}\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + \frac{2}{n_1}\mathrm{tr}\widehat{\Sigma_2^2}_{(S)}}.$$

It is easy to see that $\beta<1$ due to $n_i=N_i-1$, $i=1,2$, which leads to $T_S<T_{LC}$.

Moreover, β can be rewritten as

$$\beta = \frac{n_1n_2}{N_1N_2}\left[1+\frac{\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + \mathrm{tr}\widehat{\Sigma_2^2}_{(S)}}{n_1\,\mathrm{tr}\widehat{\Sigma_1^2}_{(S)} + n_2\,\mathrm{tr}\widehat{\Sigma_2^2}_{(S)}}\right],$$

which indicates that $\beta \to 1$ and then $T_S - T_{LC} \to 0$ as $\min\{N_1, N_2\}$ tends to infinity.

As a consequence, the fact that $T_S - T_{LC} \to 0$ as $\min\{N_1, N_2\} \to \infty$ in Theorem 4.1 also indicates that the test, proposed by Ref. [3] under the assumption (15), can be justified under the assumptions in Ref. [5], in which no relationship between p and N_i is imposed.

Finally, the numerical simulations are designed to interpret our theoretical results as well as their implications and to further evaluate the performance of T_{LC} and T_S. Let $Y_{ij} = (y_{ij1}, \cdots, y_{ijp})^T$ for $i=1, 2$ *and* $j=1, \cdots, N_i$, where $\{y_{ija}\}_{i=1,2; j=1,\cdots N_i; a=1,\cdots,p}$ are independent and identically distributed random variables with $y_{ija}+2 \sim \Gamma(4,2)$. To investigate the empirical sizes of T_{LC} and T_S, without loss of generality, we generate the first sample $\{X_{1j}\}_{j=1}^{N_1}$ by $X_{1j} = Y_{1j}$ and the second sample $\{X_{2j}\}_{j=1}^{N_2}$ by $X_{2j} = Y_{2j}$; to discuss the empirical powers of T_{LC} and T_S, we take $\{X_{1j}\}_{j=1}^{N_1}$ by $X_{1j} = Y_{1j}$ and $\{X_{2j}\}_{j=1}^{N_2}$ by $X_{2j} = \frac{9}{8} Y_{2j}$.

Tab. 2 provides the empirical sizes and powers of T_{LC} and T_S by 1000 Monte Carlo replications with $N_1 = N_2 = N$ and the nominal significance level $\alpha = 0.05$. As showed in Table 2, when N is small, T_{LC} possesses larger power than T_S, while T_S has smaller size than T_{LC}; the difference of sizes or powers between T_{LC} and T_S decreases gradually as N increases; particularly, when $N=640$, the sizes or powers of T_{LC} and T_S are almost the same. Therefore from the above statements, we may conclude that T_{LC} and T_S are asymptotically equivalent.

Tab. 2 Empirical sizes and powers of T_{LC} and T_S

p		N						N					
		20	40	80	160	320	640	20	40	80	160	320	640
		Size						Power					
20	T_{LC}	0.070	0.070	0.066	0.079	0.067	0.075	0.113	0.186	0.314	0.631	0.953	1.000
	T_S	0.061	0.067	0.062	0.078	0.067	0.075	0.100	0.168	0.312	0.629	0.953	1.000
40	T_{LC}	0.066	0.074	0.060	0.058	0.058	0.056	0.092	0.167	0.312	0.681	0.980	1.000
	T_S	0.054	0.067	0.057	0.056	0.056	0.056	0.072	0.156	0.303	0.678	0.979	1.000
80	T_{LC}	0.045	0.046	0.045	0.059	0.051	0.048	0.085	0.143	0.303	0.687	0.986	1.000
	T_S	0.036	0.041	0.044	0.058	0.051	0.048	0.067	0.132	0.296	0.683	0.985	1.000
160	T_{LC}	0.059	0.057	0.063	0.043	0.059	0.045	0.067	0.134	0.315	0.702	0.993	1.000
	T_S	0.047	0.053	0.062	0.042	0.058	0.045	0.052	0.124	0.306	0.699	0.993	1.000
320	T_{LC}	0.057	0.062	0.061	0.053	0.046	0.052	0.072	0.128	0.303	0.700	0.994	1.000
	T_S	0.042	0.054	0.060	0.053	0.046	0.052	0.061	0.118	0.299	0.697	0.994	1.000
640	T_{LC}	0.045	0.045	0.041	0.062	0.038	0.046	0.081	0.118	0.287	0.707	0.997	1.000
	T_S	0.030	0.043	0.041	0.061	0.038	0.046	0.067	0.109	0.279	0.705	0.996	1.000

5 Conclusions

Some equivalent estimators of product of two population covariance matrices Σ_1 and Σ_2 have been derived in this paper by taking the advantage that the U-statistics do not depend on the underlying distributions. This derivation is extendable and can be employed to construct the unbiased and location-invariant estimators of $\Sigma_1^m \Sigma_2^n$ and $(\Sigma_1 \Sigma_2)^m$ for any positive integers m, n. Furthermore, we find that three commonly used estimators of $\mathrm{tr}(\Sigma_1 \Sigma_2)$ respectively given in Refs. [5], [3] and [6] are exactly the same, whereupon two test statistics proposed in Ref. [5] and Ref. [3] for testing the equality of two covariance matrices are asymptotically equivalent.

References:

[1] Bai Z D, Saranadasa H. Effect of high dimension: By an example of a two sample problem [J]. Stat Sinica, 1996, 6: 311.

[2] Chen S X, Zhang L X, Zhong P S. Tests for high-dimensional covariance matrices [J]. J Am Stat Assoc, 2010, 105: 810.

[3] Srivastava M S, Yanagihara H, Kubokawa T. Tests for covariance matrices in high dimension with less sample size [J]. J Multivariate An, 2014, 130: 289.

[4] Zhou B, Guo J. A note on the unbiased estimator of Σ^2 [J]. Stat Probabil Lett, 2017, 129: 141.

[5] Li J, Chen S X. Two sample tests for high-dimensional covariance matrices [J]. Ann Stat, 2012, 40: 908.

[6] Ahmad M R. Location-invariant multi-sample U-tests for covariance matrices with large dimension [J]. Scand J Stat, 2017, 44: 500.

[7] Mao G. A note on tests for high-dimensional covariance matrices [J]. Stat Probabil Lett, 2016, 117: 89.

[8] Fisher T J. On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size [J]. J Stat Plan Infer, 2012, 142: 312.

[9] Tsuda K, Ratsch G, Warmuth M K. Matrix exponentiated gradient updates for on-line learning and Bregman projection [J]. J Mach Learn Res, 2005, 6: 995.

引用本文格式：

中　文：沈洁琼，周杰，陈晨. 协方差矩阵的乘积及其迹的估计[J]. 四川大学学报：自然科学版，2020，57：635.

英　文：Shen J Q, Zhou J, Chen C. On the estimation for product of covariance matrices and its trace [J]. J Sichuan Univ: Nat Sci Ed, 2020, 57: 635.