Journal of Multivariate Analysis 114 (2013) 349–358

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

A two sample test in high dimensional data

Muni S. Srivastava [a], Shota Katayama [b,*], Yutaka Kano [b]

[a] Department of Statistics, University of Toronto, 100 St.George Street, Toronto, Ontario M5S 3G3, Canada

[b] Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

ARTICLE INFO

Article history:
Received 6 August 2011
Available online 27 August 2012

AMS subject classifications:
primary 62H15
secondary 62E20

Keywords:
High-dimensional data
Hypothesis testing
Behrens–Fisher problem
Asymptotic theory

ABSTRACT

In this paper we propose a test for testing the equality of the mean vectors of two groups with unequal covariance matrices based on N_1 and N_2 independently distributed p-dimensional observation vectors. It will be assumed that N_1 observation vectors from the first group are normally distributed with mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$. Similarly, the N_2 observation vectors from the second group are normally distributed with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$. We propose a test for testing the hypothesis that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. This test is invariant under the group of $p \times p$ nonsingular diagonal matrices. The asymptotic distribution is obtained as $(N_1, N_2, p) \to \infty$ and $N_1/(N_1 + N_2) \to k \in (0, 1)$ but N_1/p and N_2/p may go to zero or infinity. It is compared with a recently proposed non-invariant test. It is shown that the proposed test performs the best.

1. Introduction

Let $\boldsymbol{x}_{ij}$ be independently distributed as the multivariate normal distribution with the mean vector $\boldsymbol{\mu}_i$ and the positive definite covariance matrix $\boldsymbol{\Sigma}_i$ for $i = 1, 2$ and $j = 1, 2, \ldots, N_i$. For notational convenience, we shall denote it as $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),\ i = 1, 2$, where p denotes the dimension of the random vectors $\boldsymbol{x}_{ij}$. In this article, we consider the problem of testing the hypothesis

$$H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \tag{1.1}$$

against the alternative

$$A : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2, \tag{1.2}$$

when the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ of the two groups may be unequal. This problem has recently been considered by Chen and Qin [2] who proposed a test which we denote by T_{cq}. The test T_{cq} will be described in Section 2 from which it will be clear that it is a rather complicated test and requires considerable terms in programming and computing. Also, it is shown that the T_{cq} test is almost identical to a test that can be obtained by generalizing the Bai and Saranadasa [1] test when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. In addition, the test T_{cq}, although invariant under the group of orthogonal transformations, is not invariant under the units of measurements. That is, if we consider $\boldsymbol{D}\boldsymbol{x}_{ij}$ instead of $\boldsymbol{x}_{ij}$, where $\boldsymbol{D}$ is a nonsingular $p \times p$ diagonal matrix, the test T_{cq} changes, which is an undesirable feature. It may be noted that when N_i is less than p, no fully affine invariant test exists. Thus, in this article, we propose a test that is invariant under the transformation of the observation vector $\boldsymbol{x}_{ij}$ by nonsingular

* Corresponding author.
E-mail addresses: srivasta@utstat.utoronto.ca (M.S. Srivastava), sfujimoto@sigmath.es.osaka-u.ac.jp (S. Katayama), kano@sigmath.es.osaka-u.ac.jp (Y. Kano).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.08.014

$p \times p$ diagonal matrices. It will be shown that this new test, denoted by T, performs better than T_{cq}. To describe this new test T, we introduce some notations with $n_i = N_i - 1,\ i = 1, 2$:

$$\bar{\boldsymbol{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} \boldsymbol{x}_{ij} \quad \text{and} \quad \boldsymbol{S}_i = \frac{1}{n_i}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'. \tag{1.3}$$

In high dimensional data, since N_i may be less than p, the sample covariance matrices $\boldsymbol{S}_i$ may be singular. However, the diagonal matrices consisting of only the diagonal elements of $\boldsymbol{S}_i = (s_{ijk}),\ i = 1, 2$, namely,

$$\hat{\boldsymbol{D}}_i = \operatorname{diag}(s_{i11}, \ldots, s_{ipp}), \quad i = 1, 2, \tag{1.4}$$

are non-singular matrices. Let

$$\hat{\boldsymbol{D}} = \frac{\hat{\boldsymbol{D}}_1}{N_1} + \frac{\hat{\boldsymbol{D}}_2}{N_2} = (\hat{d}_{ij}). \tag{1.5}$$

Then

$$\boldsymbol{R} = \hat{\boldsymbol{D}}^{-1/2}\left(\frac{\boldsymbol{S}_1}{N_1} + \frac{\boldsymbol{S}_2}{N_2}\right)\hat{\boldsymbol{D}}^{-1/2} = (r_{ij}) \tag{1.6}$$

is the sample correlation matrix, while $\boldsymbol{S}_i$ may not converge to $\boldsymbol{\Sigma}_i$ in probability since N_i may be less than p, $\hat{\boldsymbol{D}}_i$ converges in probability to $\boldsymbol{D}_i$, where

$$\boldsymbol{D}_i = \operatorname{diag}(\sigma_{i11}, \ldots, \sigma_{ipp}), \qquad \boldsymbol{\Sigma}_i = (\sigma_{ijk}), \quad i = 1, 2, \tag{1.7}$$

if $\max_{1\le k\le p} \sigma_{ikk} < \infty$ uniformly in p. Let

$$\boldsymbol{D} = \frac{\boldsymbol{D}_1}{N_1} + \frac{\boldsymbol{D}_2}{N_2} = (d_{ij}). \tag{1.8}$$

Then, $\hat{\boldsymbol{D}} \to \boldsymbol{D}$ in probability. Similar to the sample correlation matrix $\boldsymbol{R}$, we define the population correlation matrix $\boldsymbol{\mathcal{R}}$ by

$$\boldsymbol{\mathcal{R}} = \boldsymbol{D}^{-1/2}\left(\frac{\boldsymbol{\Sigma}_1}{N_1} + \frac{\boldsymbol{\Sigma}_2}{N_2}\right)\boldsymbol{D}^{-1/2} = (\rho_{ij}). \tag{1.9}$$

We note that under the null hypothesis H in (1.1),

$$\begin{aligned} E[(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)] &= \operatorname{tr} \boldsymbol{D}^{-1}\left(\frac{\boldsymbol{\Sigma}_1}{N_1} + \frac{\boldsymbol{\Sigma}_2}{N_2}\right) \\ &= \operatorname{tr} \boldsymbol{\mathcal{R}} = p. \end{aligned}$$

Also, under the null hypothesis H in (1.1),

$$\begin{aligned} \operatorname{Var}[(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)] &= \operatorname{Var}(\bar{\boldsymbol{x}}_1'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_2 - 2\bar{\boldsymbol{x}}_1'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_2) \\ &= \operatorname{Var}(\bar{\boldsymbol{x}}_1'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_1) + \operatorname{Var}(\bar{\boldsymbol{x}}_2'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_2) + 4\operatorname{Var}(\bar{\boldsymbol{x}}_1'\boldsymbol{D}^{-1}\bar{\boldsymbol{x}}_2) \\ &= \frac{2\operatorname{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_1)^2}{N_1^2} + \frac{2\operatorname{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_2)^2}{N_2^2} + \frac{4\operatorname{tr} \boldsymbol{D}^{-1}\boldsymbol{\Sigma}_1\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_2}{N_1N_2} \\ &= 2\operatorname{tr}\left[\left(\frac{\boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}_1\boldsymbol{D}^{-1/2}}{N_1}\right) + \left(\frac{\boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{D}^{-1/2}}{N_2}\right)\right]^2 \\ &= 2\operatorname{tr} \boldsymbol{\mathcal{R}}^2. \end{aligned}$$

Following Corollary 2.6 of [5], we have for $i = 1, 2$ and $j = 1, \ldots, p$ that $E(s_{ijj}^{-1}) = \sigma_{ijj}^{-1} + O(N_i^{-1})$. Hence, $s_{ijj}^{-1} = \sigma_{ijj}^{-1} + O_p(N_i^{-1})$. Thus,

$$\hat{\boldsymbol{D}}^{-1} = \boldsymbol{D}^{-1}[1 + O_p(N_m^{-1})], \qquad N_m = \min(N_1, N_2),$$

which implies

$$\begin{aligned} \hat{q}_n &= \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\hat{\boldsymbol{D}}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p}{\sqrt{p}} \\ &= \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p[1 + O_p(N_m^{-1})]}{\sqrt{p}}[1 + O_p(N_m^{-1})] \\ &= \tilde{q}_n + O_p\left(\frac{\sqrt{p}}{N_m}\right), \end{aligned} \tag{1.10}$$

where

$$\tilde{q}_n = \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p}{\sqrt{p}}. \tag{1.11}$$

Assume $N_m = O(p^\delta)$, $\delta > 1/2$, then it follows that $O_p(\sqrt{p}/N_m) = o_p(1)$. Thus, $\hat{q}_n \to \tilde{q}_n$ in probability. We note under the null hypothesis H in (1.1) that $E(\tilde{q}_n) = 0$ and the variance of $\tilde{q}_n$ is given by

$$\mathrm{Var}(\tilde{q}_n) = \frac{2\mathrm{tr}\,\mathcal{R}^2}{p} \simeq \mathrm{Var}(\hat{q}_n). \tag{1.12}$$

It may be noted that $E(\tilde{q}_n) = p^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq 0$, which takes the value 0 if and only if $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Thus, our proposed test is a one-sided test. Similarly, it can be shown that all other tests considered in this paper are also one-sided.

For obtaining a ratio consistent estimator of it (the ratio between it and the estimator converges to 1 in probability) and in order to find an asymptotic null distribution of $\hat{q}_n$, we make the following assumptions:

Assumption (A).

(A1) $0 < c_1 < \min_{1\leq k\leq p} \sigma_{ikk} \leq \max_{1\leq k\leq p} \sigma_{ikk} < c_2 < \infty$ uniformly in p,
(A2) $\lim_{p\to\infty} \mathrm{tr}\,\mathcal{R}^4/(\mathrm{tr}\,\mathcal{R}^2)^2 = 0$,
(A3) $N_1/N \to k \in (0, 1)$ as $N = N_1 + N_2 \to \infty$,
(A4) $N_m = O(p^\delta)$, $\delta > 1/2$, $N_m = \min(N_1, N_2)$.

Remark 1.1. Chen and Qin [2] assume the condition

$$\lim_{p\to\infty} \frac{\mathrm{tr}\,\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_\ell}{\{\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\}^2} = 0, \quad i, j, k, \ell = 1 \text{ or } 2 \tag{1.13}$$

instead of (A2). Without loss of generality, we can assume $N_1 \leq N_2$. Since

$$\frac{2c_1N}{N_1N_2}\boldsymbol{I}_p \leq \boldsymbol{D} = \frac{\boldsymbol{D}_1}{N_1} + \frac{\boldsymbol{D}_2}{N_2} \leq \frac{2c_2N}{N_1N_2}\boldsymbol{I}_p$$

under the assumption (A1), it follows that

$$\mathrm{tr}\,\mathcal{R}^4 \leq \frac{1}{N_1^4}\mathrm{tr}\{\boldsymbol{D}^{-1}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\}^4 \leq \frac{N_2^4}{16c_1^4N^4}\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^4,$$
$$\mathrm{tr}\,\mathcal{R}^2 \geq \frac{1}{N_2^2}\mathrm{tr}\{\boldsymbol{D}^{-1}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\}^2 \geq \frac{N_1^2}{4c_2^2N^2}\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2.$$

Hence,

$$\frac{\mathrm{tr}\,\mathcal{R}^4}{(\mathrm{tr}\,\mathcal{R}^2)^2} \leq \left(\frac{c_2N_2}{c_1N_1}\right)^4 \frac{\mathrm{tr}\,\boldsymbol{\Sigma}_1^4 + 4\mathrm{tr}\,\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2 + 6\mathrm{tr}\,\boldsymbol{\Sigma}_1^2\boldsymbol{\Sigma}_2^2 + 4\mathrm{tr}\,\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^3 + \mathrm{tr}\,\boldsymbol{\Sigma}_2^4}{\{\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}^2},$$

which implies that (A2) is weaker than (1.13). We note that the assumption (A1) is not required in [2]. However, the assumption (A1) is not so strong and unrealistic.

It may be noted that

$$\begin{aligned}\frac{1}{p}\mathrm{tr}\,\mathcal{R}^2 &= \frac{1}{p}\left[\mathrm{tr}\left\{\boldsymbol{D}^{-1}\left(\frac{\boldsymbol{\Sigma}_1}{N_1} + \frac{\boldsymbol{\Sigma}_2}{N_2}\right)\right\}^2\right] \\ &= \frac{1}{pN_1^2}\mathrm{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_1)^2 + \frac{1}{pN_2^2}\mathrm{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_2)^2 + \frac{2}{pN_1N_2}\mathrm{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_1\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_2) \\ &= F_1 + F_2 + 2G, \quad \text{say,}\end{aligned} \tag{1.14}$$

where

$$F_i = \frac{1}{pN_i^2}\mathrm{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_i)^2, \qquad G = \frac{1}{pN_1N_2}\mathrm{tr}(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_1\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_2). \tag{1.15}$$

From Lemma 2.1 of [4] and Lemma 3.2 of [6], under Assumption (A), it is found that F_i and G can be estimated by ratio consistent estimators

$$\hat{F}_i = \frac{1}{p}[\mathrm{tr}(\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_i)^2 - n_i^{-1}(\mathrm{tr}\,\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_i)^2], \tag{1.16}$$

and

$$\hat{G} = \frac{1}{p}\operatorname{tr}(\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_1\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_2), \tag{1.17}$$

respectively. Thus, a ratio consistent estimator of $\operatorname{Var}(\hat{q}_n)$ under Assumption (A) is given by

$$\begin{aligned}\widehat{\operatorname{Var}}(\hat{q}_n) &= \frac{2\hat{F}_1}{N_1^2} + \frac{2\hat{F}_2}{N_2^2} + \frac{4\hat{G}}{N_1N_2} \\ &= \frac{2\operatorname{tr}\boldsymbol{R}^2}{p} - \frac{2}{pn_1N_1^2}(\operatorname{tr}\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_1)^2 - \frac{2}{pn_2N_2^2}(\operatorname{tr}\hat{\boldsymbol{D}}^{-1}\boldsymbol{S}_2)^2.\end{aligned} \tag{1.18}$$

For testing the null hypothesis H in (1.1), we propose the statistic

$$T = \frac{\hat{q}_n}{\sqrt{\widehat{\operatorname{Var}}(\hat{q}_n)c_{p,n}}} = \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\hat{\boldsymbol{D}}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p}{\sqrt{p\widehat{\operatorname{Var}}(\hat{q}_n)c_{p,n}}}, \tag{1.19}$$

where

$$c_{p,n} = 1 + \frac{\operatorname{tr}\boldsymbol{R}^2}{p^{3/2}}. \tag{1.20}$$

This $c_{p,n}$ is a correction term going to one to speed up the convergence. This correction term has been given by Srivastava and Du [6] in the connection of a test when the covariance matrices of the two groups are equal. See [5] for the robustness of Srivastava and Du [6] test.

For the asymptotic distribution of the statistic T under the hypothesis H in (1.1), we have the following theorem:

Theorem 1.1. *Under the null hypothesis H in (1.1) and under the Assumption (A), the asymptotic distribution of the statistic T as $(N, p) \to \infty$ is given by*

$$\lim_{(N,p)\to\infty} P_0(T > z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha}),$$

where Φ denotes the distribution function of a standard normal random variable with mean 0 and variance 1, $z_{1-\alpha}$ is the upper $(1-\alpha) \times 100\%$ point of the standard normal distribution, and P_0 denotes that the distribution have been obtained under the hypothesis that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

To obtain the asymptotic distribution of T under the alternative in (1.2), we choose the local alternative as defined in the following assumption:

Assumption (B). As $(N, p) \to \infty$

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1/2}\mathcal{R}\boldsymbol{D}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(\operatorname{tr}\mathcal{R}^2).$$

Remark 1.2. Instead of Assumption (B), Chen and Qin [2] assume

$$N(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o\{\operatorname{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}, \quad i = 1, 2,$$

which is clearly stronger when $\boldsymbol{D}_i = \sigma_i\boldsymbol{I}_p$ for $i = 1, 2$.

The next theorem gives the asymptotic distribution of the test statistic T under the local alternative defined in Assumption (B).

Theorem 1.2. *Under Assumptions (A) and (B), the asymptotic distribution of the statistic T is given by*

$$\lim_{(N,p)\to\infty} P_1(T > z_{1-\alpha}) = \Phi\left(-z_{1-\alpha} + \frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{2\operatorname{tr}\mathcal{R}^2}}\right),$$

where P_1 denotes that the distribution has been obtained under the alternative in (1.2), and $z_{1-\alpha}$ is the upper $(1-\alpha) \times 100\%$ point of the statndard normal distribution.

The organization of the paper is as follows. In Section 2, we describe the statistic T_{cq} proposed by Chen and Qin [2] and compare the performance with the proposed statistic T. In Section 3, we compare the powers of the proposed statistic with T_{cq} statistic as well as with the usual generalized statistic T_u defined in Section 2 theoretically and numerically. In Section 4, we give the proofs of Theorems 1.1 and 1.2. The concluding remarks are given in Section 5.

2. Recently proposed statistic

In this section, we describe a recently proposed test statistic for testing the hypothesis that the mean vectors of the two groups are equal when the number of observations from each groups are smaller than the dimension and when it has been found that the two groups have unequal covariance matrices by the test given by, e.g., [7]. A simple generalization of the test statistic proposed by Bai and Saranadasa [1] is given by

$$Q_n = [(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - N_1^{-1}\mathrm{tr}\,\boldsymbol{S}_1 - N_2^{-1}\mathrm{tr}\,\boldsymbol{S}_2]/p^{1/2}.$$

The variance of this statistic is given by

$$\sigma_Q^2 = \frac{2}{pN_1^2}\mathrm{tr}\,\boldsymbol{\Sigma}_1^2 + \frac{2}{pN_2^2}\mathrm{tr}\,\boldsymbol{\Sigma}_2^2 + \frac{4}{pN_1N_2}\mathrm{tr}\,\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2.$$

A consistent estimator of σ_Q^2 from [5] is given by

$$\hat{\sigma}_Q^2 = \frac{2}{N_1^2}\hat{a}_{21} + \frac{2}{N_2^2}\hat{a}_{22} + \frac{4}{pN_1N_2}\mathrm{tr}\,\boldsymbol{S}_1\boldsymbol{S}_2,$$

where

$$\begin{aligned}\hat{a}_{2i} &= \frac{n_i^2}{p(n_i-1)(n_i+1)}\left[\mathrm{tr}\,\boldsymbol{S}_i^2 - \frac{1}{n_i}(\mathrm{tr}\,\boldsymbol{S}_i)^2\right]\\ &\simeq \frac{1}{p}\left[\mathrm{tr}\,\boldsymbol{S}_i^2 - \frac{1}{n_i}(\mathrm{tr}\,\boldsymbol{S}_i)^2\right], \quad i = 1, 2.\end{aligned}$$

Under the normality assumption $\hat{a}_{2i}$ is the best estimator of tr $\boldsymbol{\Sigma}_i^2/p$ in the sense that it is unbiased and is based on sufficient statistics and thus will have uniformly minimum variance among all unbiased estimators. Thus, the usual test statistic for testing the null hypothesis H in (1.1) can be based on

$$T_u = \frac{Q_n}{\hat{\sigma}_Q}.$$

Before describing the test statistic T_{cq} proposed by Chen and Qin [2], we observe that

$$(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \frac{1}{N_1}\mathrm{tr}\,\boldsymbol{S}_1 - \frac{1}{N_2}\mathrm{tr}\,\boldsymbol{S}_2 = \left(\bar{\boldsymbol{x}}_1'\bar{\boldsymbol{x}}_1 - \frac{1}{N_1}\mathrm{tr}\,\boldsymbol{S}_1\right) + \left(\bar{\boldsymbol{x}}_2'\bar{\boldsymbol{x}}_2 - \frac{1}{N_2}\mathrm{tr}\,\boldsymbol{S}_2\right) - 2\bar{\boldsymbol{x}}_1'\bar{\boldsymbol{x}}_2,$$

and since

$$n_i\mathrm{tr}\,\boldsymbol{S}_i = \sum_{k=1}^{N_i}(\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i)'(\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_i) = \sum_{k=1}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{ik} - \frac{1}{N_i}\sum_{k,\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell}, \quad i = 1, 2,$$

we have for $i = 1, 2$,

$$\begin{aligned}\bar{\boldsymbol{x}}_i'\bar{\boldsymbol{x}}_i - \frac{\mathrm{tr}\,\boldsymbol{S}_i}{N_i} &= \frac{1}{N_i^2}\sum_{k,\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell} - \frac{1}{N_in_i}\left(\sum_{k=1}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{ik} - \frac{1}{N_i}\sum_{k,\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell}\right)\\ &= \left(\frac{1}{N_i^2} + \frac{1}{N_i^2n_i}\right)\sum_{k,\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell} - \frac{1}{N_in_i}\sum_{k=1}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{ik}\\ &= \frac{1}{N_in_i}\sum_{k,\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell} - \frac{1}{N_in_i}\sum_{k=1}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{ik}\\ &= \frac{1}{N_in_i}\sum_{k\neq\ell}^{N_i}\boldsymbol{x}_{ik}'\boldsymbol{x}_{i\ell}.\end{aligned}$$

Thus, it follows that

$$(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \frac{1}{N_1}\mathrm{tr}\,\boldsymbol{S}_1 - \frac{1}{N_2}\mathrm{tr}\,\boldsymbol{S}_2 = \frac{1}{N_1n_1}\sum_{i\neq j}^{N_1}\boldsymbol{x}_{1i}'\boldsymbol{x}_{1j} + \frac{1}{N_2n_2}\sum_{i\neq j}^{N_2}\boldsymbol{x}_{2i}'\boldsymbol{x}_{2j} - 2\bar{\boldsymbol{x}}_1'\bar{\boldsymbol{x}}_2.$$

Hence, the statistic T_{cq} proposed by Chen and Qin [2] is given by

$$T_{cq} = \frac{Q_n}{\tilde{\sigma}_Q},$$

where

$$\tilde{\sigma}_Q^2 = \frac{1}{p}\left[\frac{2}{N_1 n_1}\widehat{\mathrm{tr}\,\boldsymbol{\Sigma}_1^2} + \frac{2}{N_2 n_2}\widehat{\mathrm{tr}\,\boldsymbol{\Sigma}_2^2} + \frac{4}{N_1 N_2}\widehat{\mathrm{tr}\,\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2}\right],$$

$$\widehat{\mathrm{tr}\,\boldsymbol{\Sigma}_i^2} = \frac{1}{N_i n_i}\mathrm{tr}\left\{\sum_{j\neq k}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i(j,k)})\boldsymbol{x}'_{ij}(\boldsymbol{x}_{ik} - \bar{\boldsymbol{x}}_{i(j,k)})\boldsymbol{x}'_{ik}\right\}, \quad i = 1, 2,$$

$$\widehat{\mathrm{tr}\,\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2} = \frac{1}{N_1 N_2}\mathrm{tr}\left\{\sum_{j=1}^{N_1}\sum_{k=1}^{N_2}(\boldsymbol{x}_{1j} - \bar{\boldsymbol{x}}_{1(j)})\boldsymbol{x}'_{1j}(\boldsymbol{x}_{2k} - \bar{\boldsymbol{x}}_{2(k)})\boldsymbol{x}'_{2k}\right\},$$

with

$$\bar{\boldsymbol{x}}_{i(j,k)} = \frac{1}{N_i - 2}(N_i\bar{\boldsymbol{x}}_i - \boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}), \quad i = 1, 2;\ j, k = 1, \ldots, N_i,$$

$$\bar{\boldsymbol{x}}_{i(k)} = \frac{1}{n_i}(N_i\bar{\boldsymbol{x}}_i - \boldsymbol{x}_{ik}), \quad i = 1, 2;\ k = 1, \ldots, N_i.$$

Clearly, the statistic T_{cq} differs from the statistic T_u in that it uses a different estimator of the variance σ_Q^2. No theoretical reasons have been given as to why one should use their estimator of σ_Q^2 than the uniformly minimum variance unbiased estimator of σ_Q^2 in the case where the observations are normally distributed. Indeed it is shown that the performance of the usual test statistic T_u is no inferior than the test statistic T_{cq}. In addition, the estimator $\tilde{\sigma}_Q$ requires much more computing and programming and takes considerably longer time in simulation. Irrespective of the selection between T_u and T_{cq}, the proposed statistic performs better than both of them. The asymptotic distribution of the statistic T_u can be obtained on the same lines as the one obtained for the statistic T. It may be noted that Chen and Qin [2] have obtained the distribution of T_{cq} under weaker conditions than normality.

3. Power comparison

In this section we compare the power of the proposed test with that of the test T_{cq} since the asymptotic power of the test T_u is identical to that of T_{cq} as both tests have the same numerator in the statistic while the denominator converges to the same quantity σ_Q^2. We first do a theoretical comparison in Section 3.1, and then in Section 3.2, we compare them by simulation.

3.1. Theoretical power comparison

The theoretical power of the proposed test for large (N, p) is given in Theorem 1.2 of Section 1. The theoretical power of T_{cq} (as well as of T_u) has been derived by Chen and Qin [2]. It is given by

$$\lim_{(N,p)\to\infty} P_1(T_{cq} > z_{1-\alpha}) = \Phi\left[-z_{1-\alpha} + \frac{Nk(1-k)\boldsymbol{\delta}'\boldsymbol{\delta}}{\sqrt{2\mathrm{tr}\,\tilde{\boldsymbol{\Sigma}}(k)^2}}\right], \tag{3.1}$$

where $\tilde{\boldsymbol{\Sigma}}(k) = (1-k)\boldsymbol{\Sigma}_1 + k\boldsymbol{\Sigma}_2,\ k = N_1/N, N = N_1 + N_2$ and $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. It may be noted that

$$\begin{aligned}[Nk(1-k)]^{-2}\mathrm{tr}\,\tilde{\boldsymbol{\Sigma}}(k)^2 &= N_1^{-2}\mathrm{tr}\,\boldsymbol{\Sigma}_1^2 + N_2^{-2}\mathrm{tr}\,\boldsymbol{\Sigma}_2^2 + 2(N_1N_2)^{-1}\mathrm{tr}\,\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\\ &= n^{*-2}\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2,\end{aligned}$$

when $N_1 = N_2 = n^*$, that is, when the sample sizes are equal. Furthermore, for diagonal matrices $\boldsymbol{\Sigma}_i = \boldsymbol{D}_i = \mathrm{diag}(d_{i1}, \ldots, d_{ip})$, the power given in (3.1) becomes

$$\Phi\left[-z_{1-\alpha} + \frac{n^*\boldsymbol{\delta}'\boldsymbol{\delta}}{\sqrt{2c}}\right], \tag{3.2}$$

where $c = \sum_{j=1}^p c_j^2$ with $c_j = d_{1j} + d_{2j}$. The power of the proposed test for equal sample size n^* and diagonal matrices for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ becomes

$$\Phi\left[-z_{1-\alpha} + \frac{n^*\boldsymbol{\delta}'\boldsymbol{D}_c^{-1}\boldsymbol{\delta}}{\sqrt{2p}}\right] \tag{3.3}$$

since $\boldsymbol{\mathcal{R}} = \boldsymbol{I}_p$. Here $\boldsymbol{D}_c = \boldsymbol{D}_1 + \boldsymbol{D}_2 = \mathrm{diag}(c_1, \ldots, c_p)$.

It may be noted that since the test statistic T_{cq} is invariant under the group of orthogonal transformations, the two covariance matrices can be assumed to be diagonal for equal sample sizes provided $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are exchangeable, namely,

$\mathbf{\Sigma}_1\mathbf{\Sigma}_2 = \mathbf{\Sigma}_2\mathbf{\Sigma}_1$. For theoretical comparison, we make this assumption. It may be noted that the proposed statistic T, although invariant under nonsingular diagonal matrices, it is not orthogonally invariant. Furthermore, we assume that the local alternatives for both tests are such that

$$0 < \lim_{p\to\infty} \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p} = \lim_{p\to\infty} \frac{\boldsymbol{\delta}'\boldsymbol{D}_c^{-1}\boldsymbol{\delta}}{\operatorname{tr} \boldsymbol{D}_c^{-1}} < \infty. \tag{3.4}$$

It is noted that (3.4) is satisfied if $\boldsymbol{\delta} = (\delta, \dots, \delta)'$, $\delta \neq 0$. It is shown in Eq. (6.3) of [6] that

$$\frac{1}{p}\sum_{i=1}^{p}\frac{1}{c_i} \geq \left(\frac{p}{\sum_{i=1}^{p} c_i^2}\right)^{1/2}, \tag{3.5}$$

and hence for $\boldsymbol{\mathcal{R}} = \boldsymbol{I}_p$,

$$\begin{aligned}\frac{\boldsymbol{\delta}'\boldsymbol{D}_c^{-1}\boldsymbol{\delta}}{\sqrt{\operatorname{tr} \boldsymbol{\mathcal{R}}^2}} &= \frac{\boldsymbol{\delta}'\boldsymbol{D}_c^{-1}\boldsymbol{\delta}}{\sqrt{p}} = \frac{\boldsymbol{\delta}'\boldsymbol{D}_c^{-1}\boldsymbol{\delta}}{\sum_{i=1}^{p} c_i^{-1}}\frac{\sum_{i=1}^{p} c_i^{-1}}{\sqrt{p}} \\ &\simeq \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p}\frac{\sum_{i=1}^{p} c_i^{-1}}{\sqrt{p}} \geq \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\sqrt{\sum_{i=1}^{p} c_i^2}} = \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\sqrt{c}},\end{aligned} \tag{3.6}$$

where '$\geq$' is strict unless $c_1 = \cdots = c_p$. Thus, we get the following theorem:

Theorem 3.1. *Assume that $\mathbf{\Sigma}_i = \operatorname{diag}(d_{i1}, \dots, d_{ip})$, $N_1 = N_2 = n^*$ and the Assumption (A). Under the local alternatives satisfying (3.4), the power denoted by β of the three tests have the following relationship:*

$$\beta(T|\boldsymbol{\delta}) \geq \beta(T_{cq}|\boldsymbol{\delta}) = \beta(T_u|\boldsymbol{\delta}),$$

with strict inequality unless $c_1 = \cdots = c_p$.

Thus, in the case of equal sample sizes $N_1 = N_2$, the proposed test is superior to the tests using the statistics T_{cq} and T_u.

3.2. Comparison of powers by simulation

In this section, we compare the performance of the proposed statistic T with the usual statistic T_u and T_{cq} in finite samples and dimensions by simulation. We first examine the attained significance level (ASL) of the test statistics T and T_{cq} compared to the nominal value $\alpha = 0.05$, and then we examine their attained power. Assume that the data is generated from the model

$$\boldsymbol{x}_{ij} = \boldsymbol{\mu}_i + \mathbf{\Sigma}_i^{1/2}\boldsymbol{z}_{ij}, \quad i = 1, 2;\ j = 1, \dots, n_i,$$

where $\boldsymbol{z}_{ij} = (z_{ij1}, \dots, z_{ijp})^T$ and z_{ijk}'s are independent random variables which are distributed as either of the following three distributions:

(i) $N(0, 1)$, (ii) $(\chi_2^2 - 2)/4$, (iii) $(\chi_8^2 - 8)/4$.

The ASL is computed as $\hat{\alpha} = \#(T_H > z_{1-\alpha})/r$ where T_H are values of the test statistic obtained from data simulated under H, r is the number of replications and $z_{1-\alpha}$ is the $100(1-\alpha)\%$ point of the standard normal distribution. From this simulation, we also obtain $\hat{z}_{1-\alpha}$ as the $100(1-\alpha)\%$ point of the empirical distribution of T_H. We define the attained power of the test T as $\hat{\beta} = \#(T_A > \hat{z}_{1-\alpha})/r$, where T_A are values of the test statistic computed from data simulated under the alternative.

In Tables 1 and 2, the ASL and the attained power of T, T_{cq} and T_u are given for $\mathbf{\Sigma}_1 = \boldsymbol{D}\boldsymbol{\mathcal{R}}_1\boldsymbol{D}$, $\boldsymbol{D} = \operatorname{diag}(d_1, \dots, d_p)$, $d_i = 2 + (p-i+1)/p$, $\boldsymbol{\mathcal{R}}_1 = (r_{ij})$, $r_{ii} = 1$, $r_{ij} = (-1)^{i+j}(0.2)^{|i-j|^{0.1}}$, $i \neq j$ and $\mathbf{\Sigma}_2 = \mathbf{\Psi}\boldsymbol{\mathcal{R}}_2\mathbf{\Psi}$, $\mathbf{\Psi} = \operatorname{diag}(\psi_1, \dots, \psi_p)$ with $\psi_i = 4 + (p-i+1)/p$, $\boldsymbol{\mathcal{R}}_2 = (\rho_{ij})$, $\rho_{ii} = 1$, $\rho_{ij} = (-1)^{i+j}(0.4)^{|i-j|^{0.1}}$, $i \neq j$. For the null hypothesis, we choose $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$ and for the alternative we choose $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = (u_1, \dots, u_p)'$, where $u_i = (-1)^i v_i$ with v_i are i.i.d. as $U(1/2, 3/2)$ which denotes uniform distribution with the support $(1/2, 3/2)$. In Tables 3 and 4, the ASL and the power are given when ψ_i i.i.d. as χ_3^2 with the rest remaining the same. In the tables, we use 1000 replications of the test statistic. Also we take $N_1 = N_2 = n^*$ for simplicity.

Table 1
Attained significance levels of T, T_{cq} and T_u when $\psi_i = 4 + (p - i + 1)/p$ and $n^* = N_1 = N_2$.

p	n^*	z_{ijk} i.i.d. $N(0, 1)$			z_{ijk} i.i.d. $(\chi_2^2 - 2)/2$			z_{ijk} i.i.d. $(\chi_8^2 - 8)/4$		
		T	T_{cq}	T_u	T	T_{cq}	T_u	T	T_{cq}	T_u
60	30	0.083	0.090	0.090	0.072	0.072	0.069	0.073	0.078	0.079
100	40	0.074	0.091	0.090	0.071	0.073	0.073	0.084	0.097	0.097
	60	0.066	0.086	0.086	0.049	0.066	0.065	0.057	0.066	0.067
	80	0.054	0.069	0.069	0.071	0.084	0.083	0.052	0.074	0.074
	100	0.054	0.082	0.082	0.043	0.065	0.065	0.049	0.075	0.075
150	40	0.062	0.072	0.075	0.068	0.082	0.081	0.058	0.072	0.072
	60	0.057	0.079	0.079	0.054	0.067	0.066	0.059	0.075	0.077
	80	0.063	0.091	0.091	0.058	0.079	0.079	0.048	0.067	0.067
	100	0.049	0.073	0.074	0.058	0.070	0.070	0.055	0.067	0.066
	150	0.055	0.079	0.078	0.045	0.066	0.066	0.041	0.050	0.050
	200	0.045	0.066	0.066	0.042	0.057	0.057	0.053	0.073	0.073
200	40	0.055	0.068	0.068	0.058	0.077	0.073	0.067	0.084	0.081
	60	0.060	0.078	0.079	0.057	0.074	0.073	0.072	0.093	0.093
	80	0.067	0.087	0.085	0.044	0.069	0.069	0.045	0.065	0.065
	100	0.060	0.087	0.088	0.041	0.068	0.067	0.047	0.062	0.062
	150	0.043	0.069	0.070	0.048	0.076	0.075	0.044	0.063	0.063
	200	0.043	0.076	0.076	0.042	0.068	0.068	0.048	0.067	0.067
	250	0.045	0.063	0.063	0.047	0.072	0.071	0.053	0.068	0.070

Table 2
Attained powers of T, T_{cq} and T_u when $\psi_i = 4 + (p - i + 1)/p$ and $n^* = N_1 = N_2$.

p	n^*	z_{ijk} i.i.d. $N(0, 1)$			z_{ijk} i.i.d. $(\chi_2^2 - 2)/2$			z_{ijk} i.i.d. $(\chi_8^2 - 8)/4$		
		T	T_{cq}	T_u	T	T_{cq}	T_u	T	T_{cq}	T_u
60	30	0.245	0.249	0.250	0.316	0.327	0.324	0.320	0.335	0.338
100	40	0.378	0.368	0.371	0.384	0.403	0.407	0.386	0.409	0.413
	60	0.537	0.519	0.521	0.590	0.552	0.558	0.616	0.610	0.621
	80	0.664	0.673	0.670	0.678	0.669	0.674	0.678	0.680	0.688
	100	0.810	0.814	0.814	0.842	0.821	0.822	0.834	0.819	0.821
150	40	0.456	0.468	0.466	0.435	0.442	0.433	0.430	0.438	0.435
	60	0.587	0.600	0.601	0.632	0.631	0.632	0.625	0.632	0.636
	80	0.672	0.670	0.668	0.733	0.730	0.731	0.706	0.712	0.713
	100	0.823	0.828	0.830	0.833	0.830	0.829	0.810	0.807	0.807
	150	0.964	0.964	0.964	0.962	0.955	0.955	0.966	0.960	0.960
	200	0.995	0.993	0.993	0.990	0.990	0.990	0.995	0.995	0.995
200	40	0.488	0.475	0.476	0.462	0.453	0.451	0.457	0.464	0.458
	60	0.608	0.619	0.619	0.615	0.618	0.620	0.607	0.618	0.623
	80	0.770	0.773	0.774	0.780	0.771	0.772	0.778	0.776	0.779
	100	0.808	0.792	0.794	0.896	0.894	0.893	0.896	0.893	0.893
	150	0.960	0.959	0.959	0.966	0.964	0.964	0.964	0.960	0.960
	200	0.990	0.986	0.986	0.983	0.984	0.984	0.990	0.988	0.988
	250	0.999	0.999	0.999	0.998	0.998	0.998	1.000	1.000	1.000

It is shown that the attained significance levels of the proposed test T approximate $\alpha = 0.05$ well except when the sample size is very small. As shown in Table 2, the powers of T, T_{cq} and T_u are almost the same. In Table 4, however, the power of T is substantially higher than those of T_{cq} and T_u. The reason can be given as follows. Since the non-diagonal terms of $\boldsymbol{\mathcal{R}}_1$ and $\boldsymbol{\mathcal{R}}_2$ are close to 0, it can be regarded that both $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the diagonal matrices. Then, the asymptotic powers of T and T_{cq} (or T_u) are given by (3.2) and (3.3) respectively, and the value of (3.3) is significantly larger than that of (3.2) in the settings of Table 4. Generally, the value of (3.3) seems to be larger than that of (3.2) when the difference between the maximum and minimum values of $\{\sigma_{ijj} | \, i = 1, 2; \; j = 1, \ldots, p\}$ is large.

4. Proofs of theorems in Section 1

In this section, we give the proofs of the two theorems stated in Section 1. We begin with the proof of Theorem 1.1.

4.1. Proof of Theorem 1.1

Since $\hat{q}_n \to \tilde{q}_n$ in probability as shown in Section 1, we need only to find the distribution of the statistic

$$\tilde{T} = \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{D}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p}{(2 \mathrm{tr}\, \boldsymbol{\mathcal{R}}^2)^{1/2}}.$$

Table 3
Attained significance levels of T, T_{cq} and T_u when ψ_i i.i.d. as χ^2_3 and $n^* = N_1 = N_2$.

p	n^*	z_{ijk} i.i.d. $N(0, 1)$			z_{ijk} i.i.d. $(\chi^2_2 - 2)/2$			z_{ijk} i.i.d. $(\chi^2_8 - 8)/4$		
		T	T_{cq}	T_u	T	T_{cq}	T_u	T	T_{cq}	T_u
60	30	0.077	0.085	0.089	0.117	0.084	0.072	0.091	0.069	0.068
100	40	0.069	0.089	0.089	0.117	0.097	0.092	0.078	0.069	0.069
	60	0.067	0.075	0.075	0.074	0.075	0.071	0.067	0.081	0.080
	80	0.059	0.072	0.071	0.055	0.059	0.055	0.059	0.071	0.071
	100	0.063	0.080	0.079	0.074	0.075	0.074	0.061	0.067	0.066
150	40	0.054	0.063	0.063	0.108	0.088	0.088	0.068	0.079	0.079
	60	0.058	0.061	0.062	0.074	0.082	0.081	0.073	0.074	0.072
	80	0.052	0.061	0.061	0.079	0.071	0.069	0.059	0.077	0.077
	100	0.058	0.067	0.067	0.067	0.076	0.074	0.071	0.077	0.077
	150	0.048	0.069	0.069	0.070	0.084	0.084	0.070	0.073	0.073
	200	0.066	0.080	0.080	0.055	0.065	0.065	0.046	0.059	0.059
200	40	0.073	0.079	0.080	0.113	0.093	0.089	0.085	0.086	0.085
	60	0.061	0.073	0.074	0.072	0.070	0.066	0.078	0.082	0.082
	80	0.069	0.086	0.086	0.084	0.072	0.071	0.073	0.082	0.082
	100	0.070	0.083	0.083	0.066	0.062	0.060	0.064	0.073	0.072
	150	0.057	0.085	0.085	0.054	0.073	0.069	0.039	0.047	0.045
	200	0.043	0.053	0.052	0.065	0.081	0.080	0.049	0.072	0.072
	250	0.068	0.083	0.083	0.059	0.076	0.075	0.059	0.070	0.069

Table 4
Attained powers of T, T_{cq} and T_u when ψ_i i.i.d. as χ^2_3 and $n^* = N_1 = N_2$.

p	n^*	z_{ijk} i.i.d. $N(0, 1)$			z_{ijk} i.i.d. $(\chi^2_2 - 2)/2$			z_{ijk} i.i.d. $(\chi^2_8 - 8)/4$		
		T	T_{cq}	T_u	T	T_{cq}	T_u	T	T_{cq}	T_u
60	30	0.738	0.268	0.272	0.726	0.540	0.541	0.708	0.431	0.441
100	40	0.824	0.395	0.396	0.754	0.427	0.433	0.832	0.369	0.378
	60	0.979	0.727	0.731	0.993	0.884	0.883	0.993	0.803	0.808
	80	0.998	0.884	0.883	0.999	0.959	0.960	0.998	0.953	0.956
	100	0.999	0.927	0.926	1.000	0.952	0.952	1.000	0.990	0.990
150	40	0.966	0.621	0.622	0.840	0.472	0.458	0.875	0.495	0.496
	60	0.982	0.654	0.655	0.988	0.603	0.598	0.996	0.815	0.817
	80	1.000	0.964	0.963	1.000	0.944	0.944	0.999	0.870	0.874
	100	1.000	0.978	0.979	1.000	0.858	0.853	1.000	0.982	0.982
	150	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	40	0.928	0.628	0.624	0.914	0.557	0.566	0.926	0.528	0.540
	60	0.997	0.852	0.854	0.998	0.868	0.867	0.975	0.660	0.660
	80	1.000	0.858	0.856	1.000	0.966	0.966	1.000	0.938	0.940
	100	1.000	0.960	0.960	1.000	0.996	0.996	1.000	0.986	0.987
	150	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Under normality assumption,

$$\boldsymbol{u} = \boldsymbol{\mathcal{R}}^{-1/2}\boldsymbol{D}^{-1/2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p).$$

Thus,

$$\tilde{T} = \frac{\boldsymbol{u}'\boldsymbol{\mathcal{R}}\boldsymbol{u} - p}{(2\text{tr}\,\boldsymbol{\mathcal{R}}^2)^{1/2}}.$$

Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $\boldsymbol{\mathcal{R}} = \boldsymbol{G}\boldsymbol{D}_\lambda\boldsymbol{G}'$ where $\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{I}_p$ and $\boldsymbol{D}_\lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$. Then,

$$\tilde{T} = \frac{\sum_{i=1}^{p} \lambda_i(v_i^2 - 1)}{(2\text{tr}\,\boldsymbol{\mathcal{R}}^2)^{1/2}} = \frac{\sum_{i=1}^{p} z_i}{(2\text{tr}\,\boldsymbol{\mathcal{R}}^2)^{1/2}},$$

where $\boldsymbol{v} = (v_1, \ldots, v_p)' = \boldsymbol{G}'\boldsymbol{u} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $z_i = \lambda_i(v_i^2 - 1)$. Thus, $z_1, \ldots, z_p$ are independent random variables with $E(z_i) = 0$, $E(z_i^2) = 2\lambda_i^2$ and $E(z_i^4) = 60\lambda_i^4$. Under the assumption (A2), Lyapunov's condition (see, [3, p. 332]) is satisfied for $\delta = 2$, namely,

$$\frac{\sum_{i=1}^{p} E(z_i^4)}{(2\text{tr}\,\boldsymbol{\mathcal{R}}^2)^2} = 15\frac{\text{tr}\,\boldsymbol{\mathcal{R}}^4}{(\text{tr}\,\boldsymbol{\mathcal{R}}^2)^2} \to 0, \quad p \to \infty.$$

Hence, from Lyapunov's central limit theorem, it follows that T is asymptotically normally distributed under the hypothesis that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. This proves the theorem.

4.2. *Proof of Theorem 1.2*

We note that

$$\begin{aligned}(2\mathrm{tr}\, \mathcal{R}^2)^{1/2}\tilde{T} &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\\ &\quad + 2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\end{aligned}$$

and

$$\mathrm{Var}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)] = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1/2}\mathcal{R}\boldsymbol{D}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(\mathrm{tr}\, \mathcal{R}^2).$$

Hence,

$$\frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(2\mathrm{tr}\, \mathcal{R}^2)^{1/2}} \to 0,$$

in probability from Assumption (B). From Theorem 1.1, the first term on the right hand side is asymptotically distributed as $N(0, 1)$. This ends the proof.

5. Concluding remarks

In this paper a new test statistic is proposed for testing the equality of the two mean vectors when the covariance matrices of the two groups are not equal. It is required that both $N_i = O(p^{\delta}),\ \ \delta > 1/2$ and both N_i and p go to infinity. It has been shown that the proposed test which is invariant under nonsingular diagonal matrix transformation of the observation vectors performs much better than the T_{cq} statistic proposed by Chen and Qin [2], as well as the usual test statistic which is simple and easier to compute. This usual test T_u performs as good as T_{cq}.

Acknowledgments

We would like to thank the associate editor and two referees for their careful reading of the original article and for their valuable suggestions that greatly helped improve the presentation. The research of Srivastava was supported by the Natural Sciences and Engineering Research Council of Canada, Katayama was supported by the Grant-in-Aid for JSPS Fellows and Kano was supported by the Grant-in-Aid for Challenging Exploratory Research and Grant-in-Aid for Scientific Research (B).

References

[1] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statistica Sinica 6 (1996) 311–329.
[2] S.X. Chen, Y. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, The Annals of Statistics 38 (2) (2010) 808–835.
[3] A.N. Shiryaev, Probability, second ed., Springer-Verlag, New York, 1995.
[4] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, Journal of the Japan Statistical Society 35 (2005) 251–272.
[5] M.S. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, Journal of Multivariate Analysis 100 (2009) 518–532.
[6] M.S. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, Journal of Multivariate Analysis 99 (2008) 386–402.
[7] M.S. Srivastava, H. Yanagihara, Testing the equality of several covariance matrices with fewer observations than the dimension, Journal of Multivariate Analysis 101 (2010) 1319–1329.