Statistics and Probability Letters 105 (2015) 29–36

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A note on high-dimensional two-sample test

Long Feng, Fasheng Sun *

Department of Statistics, KLAS and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 23 May 2015
Accepted 24 May 2015
Available online 9 June 2015

MSC:
62H15

Keywords:
Behrens–Fisher problem
Large p, small n
Scalar-invariant

ABSTRACT

We propose a new scalar and shift transform invariant test statistic for the high-dimensional two-sample location problem. Theoretical results and simulation studies show the good performance of our test under certain circumstances.

1. Introduction

This article is concerned with the two-sample Behrens–Fisher problem in high-dimensional settings. Assume that $\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}\}$ for $i = 1$ and 2 are two independent random samples with sizes n_1 and n_2 from p-variate distributions $F(\mathbf{x} - \boldsymbol{\mu}_1)$ and $G(\mathbf{x} - \boldsymbol{\mu}_2)$ located at the p-variate centers $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively. We wish to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2, \tag{1}$$

where their covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are unknown. If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the classic Hotelling's T^2 test is a natural choice when the dimensions are fixed. However, when the dimension is larger than the total sample size $n = n_1 + n_2$, Hotelling's T^2 test does not work. Recently, many efforts have been devoted to construct new test procedures for high-dimensional settings. One natural method involves replacing the sample covariance matrix by the identity matrix (Bai and Saranadasa, 1996; Chen and Qin, 2010; Paindaveine and Verdebout, 2013; Ley et al., 2015). However, those test statistics are not invariant under scalar transformations, $\mathbf{X}_{ij} \to \mathbf{B}\mathbf{X}_{ij}$ where $\mathbf{B}$ is a diagonal matrix. Srivastava and Du (2008) proposed a scalar-transformation-invariant test by replacing the sample covariance matrix with its diagonal matrix. And Srivastava et al. (2013) extended this method to the unequal covariance case. Feng et al. (in press) proposed another scalar-transformation-invariant test that allows the dimension with a smaller order of n^3. Gregory et al. (in press) proposed a generalized component test with $p = o(n^6)$. However, the requirement of p to be of the polynomial order of n is too restrictive in the "large p, small n" situation. For the conditions on p and n, some tests adopt permutations or simulations to compute p-values, such as Nettleton et al. (2008), Chang et al. (2014) and Thulin (2014). Park and Ayyala (2013) proposed a scalar-transformation-invariant test that

* Corresponding author.
E-mail address: sunfs359@nenu.edu.cn (F. Sun).

http://dx.doi.org/10.1016/j.spl.2015.05.017
0167-7152/

allows the dimensions to be arbitrarily large. However, their test is not shift-invariant. The differences between the ratio-consistent estimators of the diagonal matrix are not ignorable. Under the null hypothesis,

$$E(T_{PA}) = \sum_{k=1}^{p} \frac{2n^{-2}\mu_k^2(\sigma_{1k}^2 - \sigma_{2k}^2)^2}{(\kappa\sigma_{1k}^2 + (1-\kappa)\sigma_{2k}^2)^3}(1+o(1))$$

where $\sigma_{ij}^2, i = 1, 2, j = 1, \ldots, p$ are the variances of the variables, and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0 = (\mu_1, \ldots, \mu_p)$, $n_1/n \to \kappa$. When the variances of the two samples are not all equal and the common vector is very large, $E(T_{PA})$ is not zero even under the null hypothesis. In this case, we need another hypothesis test for the equality of the two covariance matrices (Li and Chen, 2012). To overcome this issue, we propose a novel test that is both scalar-invariant and shift-invariant. Under the null hypothesis, the expectation of our test statistic is exactly zero. There is no bias term in our test statistic. In addition, we do not require a relationship between the dimensions and the sample sizes. The dimension p can be arbitrarily large. The asymptotic normality of the proposed test can be derived under some mild conditions. We also proposed the asymptotic relative efficiency of our test with respect to Chen and Qin (2010)'s test. The simulation studies are consistent with the theoretical results.

The rest of the paper is organized as follows. In Section 2, we propose the new test statistic and establish its asymptotic normality. Simulation studies are conducted in Section 3. We provide the technical details in the Appendix.

2. Our test

We propose a new shift and scalar transformation invariant test statistic,

$$T_n = \frac{1}{n_1(n_1-1)}\frac{1}{n_2(n_2-1)}\sum_{k=1}^{p}\sum_{i\neq j}^{n_1}\sum^{n_1}\sum_{s\neq t}^{n_2}\sum^{n_2}\frac{(X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk})}{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2}$$

where $\gamma = n_1/n_2$, $\hat\sigma_{1k(i,j)}^2$ is the sample variance of $\{X_{1lk}\}_{l=1}^{n_1}$ excluding X_{1ik} and X_{1jk}. So does $\hat\sigma_{2k(s,t)}^2$. Because the numerator $(X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk})$ is independent of the denominator $\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2$,

$$\begin{aligned}E\left(\frac{(X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk})}{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2}\right) &= E((X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk}))E\left(\{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2\}^{-1}\right)\\ &= (\mu_{1k}-\mu_{2k})^2E\left(\{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2\}^{-1}\right).\end{aligned}$$

Unlike the three different estimators of $\sigma_{1k}^2 + \gamma\sigma_{2k}^2$ for the three parts of the test statistic in Park and Ayyala (2013), we use the leave-two-out sample variance for each numerator. $E\left(\{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2\}^{-1}\right)$ is exactly the same for each numerator and

$$E(T_n) = \sum_{k=1}^{p}(\mu_{1k}-\mu_{2k})^2E\left(\{\hat\sigma_{1k(i,j)}^2 + \gamma\hat\sigma_{2k(s,t)}^2\}^{-1}\right).$$

Under the null hypothesis H_0, $E(T_n)$ is exactly zero. Furthermore, under the Conditions (C1)–(C3) stated in the following,

$$E(T_n) = \|\boldsymbol{\Lambda}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\|^2 + o(\sqrt{\mathrm{var}(T_n)})$$

$$\mathrm{var}(T_n) = \left\{\frac{2}{n_1(n_1-1)}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda})^2) + \frac{2}{n_2(n_2-1)}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2) + \frac{4}{n_1n_2}\mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})\right\}(1+o(1))$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\left\{(\sigma_{11}^2+\gamma\sigma_{21}^2)^{-1/2}, \ldots, (\sigma_{1p}^2+\gamma\sigma_{2p}^2)^{-1/2}\right\}$.

The tests statistics proposed by Bai and Saranadasa (1996) and Chen and Qin (2010) are invariant under orthogonal transformations, $\mathbf{X}_{ij} \to \mathbf{P}\mathbf{X}_{ij}$ where $\mathbf{P}$ is an orthogonal matrix. In contrast, T_n is not invariant under orthogonal transformations, but it is invariant under location shifts and scalar transformations.

Proposition 1. *The T_n defined above is invariant under location shifts and scalar transformations. Here, the location shifts and scalar transformations mean*

$$\mathbf{X}_{ij} \to \mathbf{B}\mathbf{X}_{ij} + \mathbf{c} \quad \text{for } i = 1, 2,\ j = 1, \ldots, n_i,$$

where $\mathbf{c}$ is a constant vector, $\mathbf{B} = diag(b_1^2, \ldots, b_p^2)$, and $b_1^2, \ldots, b_p^2$ are non-zero constants.

Assume, similar to Bai and Saranadasa (1996) and Chen and Qin (2010), that $\mathbf{X}_{ij}$'s come from the following multivariate model:

$$\mathbf{X}_{ij} = \boldsymbol{\Gamma}_i\mathbf{z}_{ij} + \boldsymbol{\mu}_i \quad \text{for } j = 1, \ldots, n_i,\ i = 1, 2, \tag{2}$$

where each $\boldsymbol{\Gamma}_i$ is a $p\times m$ matrix for some $m\ge p$ such that $\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i^T=\boldsymbol{\Sigma}_i$, and $\{\mathbf{z}_{ij}\}_{j=1}^{n_i}$ are m-variate independent and identically distributed random vectors such that

$$\begin{aligned}&E(\mathbf{z}_i)=0,\qquad \mathrm{var}(\mathbf{z}_i)=\mathbf{I}_m,\qquad E(z_{il}^4)=3+\Delta,\quad \Delta>0,\qquad E(z_{il}^8)=m_8\in(0,\infty),\\ &E(z_{ik_1}^{\alpha_1}z_{ik_2}^{\alpha_2}\cdots z_{ikq}^{\alpha_q})=E(z_{ik_1}^{\alpha_1})E(z_{ik_2}^{\alpha_2})\cdots E(z_{ikq}^{\alpha_q}),\end{aligned}\tag{3}$$

for a positive integer q such that $\sum_{k=1}^q\alpha_k\le 8$ and $k_1\ne k_2\cdots\ne k_q$. The data structure generates a rich collection of $\mathbf{X}_i$ from $\mathbf{z}_i$ with a given covariance. Additionally, we need the following conditions as n and $p\to\infty$:

(C1) $n_1/(n_1+n_2)\to\kappa\in(0,1)$.
(C2) $\mathrm{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_i\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_j\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_l\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_h\boldsymbol{\Lambda}\right)=o(\mathrm{tr}^2\{(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2\})$ for $i,j,l,h=1$ or 2.
(C3) $(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_i\boldsymbol{\Lambda}^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=o(n^{-1}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2))$, for $i=1,2$. $((\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2))^2=o(n^{-1}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2))$.

The following theorem establishes the asymptotic null distribution of T_n.

Theorem 1. *Under Conditions* (C1)–(C3), *as p and $n\to\infty$,*

$$\frac{T_n-E(T_n)}{\sqrt{\mathrm{var}(T_n)}}\xrightarrow{\mathcal{L}}N(0,1).$$

Here we adopt the following ratio-consistent estimators of $\mathrm{var}(T_n)$ in Feng et al. (in press):

$$\hat\sigma_n^2\doteq\widehat{\mathrm{var}(T_n)}=\left\{\frac{2}{n_1(n_1-1)}\mathrm{tr}(\widehat{(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda})^2})+\frac{2}{n_2(n_2-1)}\mathrm{tr}(\widehat{(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2})+\frac{4}{n_1n_2}\mathrm{tr}(\widehat{\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_2\boldsymbol{\Lambda}})\right\}$$

where

$$\mathrm{tr}(\widehat{(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s\boldsymbol{\Lambda})^2})=\frac{1}{2P_{n_s}^4}\sum{}^{*}(\mathbf{X}_{si_1}-\mathbf{X}_{si_2})^T\mathbf{D}_{s(i_1,i_2,i_3,i_4)}^{-1}(\mathbf{X}_{si_3}-\mathbf{X}_{si_4})(\mathbf{X}_{si_3}-\mathbf{X}_{si_2})^T\mathbf{D}_{s(i_1,i_2,i_3,i_4)}^{-1}(\mathbf{X}_{si_1}-\mathbf{X}_{si_4}),$$

$s=1,2$, and

$$\mathrm{tr}(\widehat{\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_2\boldsymbol{\Lambda}})=\frac{1}{4P_{n_1}^2P_{n_2}^2}\sum_{i_1\ne i_2}^{n_1}\sum^{n_1}\sum_{i_3\ne i_4}^{n_2}\sum^{n_2}\left((\mathbf{X}_{1i_1}-\mathbf{X}_{1i_2})^T\mathbf{D}_{(i_1,i_2,i_3,i_4)}^{-1}(\mathbf{X}_{2i_3}-\mathbf{X}_{2i_4})\right)^2,$$

where

$$\begin{aligned}\mathbf{D}_{1(i_1,i_2,i_3,i_4)}&=\mathrm{diag}(\hat\sigma_{11(i_1,i_2,i_3,i_4)}^2+\gamma\hat\sigma_{21}^2,\ldots,\hat\sigma_{1p(i_1,i_2,i_3,i_4)}^2+\gamma\hat\sigma_{2p}^2),\\ \mathbf{D}_{2(i_1,i_2,i_3,i_4)}&=\mathrm{diag}(\hat\sigma_{11}^2+\gamma\hat\sigma_{21(i_1,i_2,i_3,i_4)}^2,\ldots,\hat\sigma_{1p}^2+\gamma\hat\sigma_{2p(i_1,i_2,i_3,i_4)}^2),\\ \mathbf{D}_{(i_1,i_2,i_3,i_4)}&=\mathrm{diag}(\hat\sigma_{11(i_1,i_2)}^2+\gamma\hat\sigma_{21(i_3,i_4)}^2,\ldots,\hat\sigma_{1p(i_1,i_2)}^2+\gamma\hat\sigma_{2p(i_3,i_4)}^2),\end{aligned}$$

and $\hat\sigma_{sk(i_1,\ldots,i_l)}^2$ is the sth sample variance after excluding X_{si_j}, $j=1,\ldots,l$, $s=1,2$, $l=2,4$, $k=1,\ldots,p$. Throughout this article, we use $\sum^*$ to denote summations over distinct indexes. For example, in $\mathrm{tr}(\widehat{(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda})^2})$, the summation is over the set $\{i_1\ne i_2\ne i_3\ne i_4\}$, for all $i_1,i_2,i_3,i_4\in\{1,\ldots,n_1\}$ and $P_n^m=n!/(n-m)!$.

This result suggests rejecting H_0 with α level of significance if $T_n/\hat\sigma_n>z_\alpha$, where z_α is the upper α quantile of $N(0,1)$. Theorem 1 shows that the power of our test is

$$\beta_{FS}(\|\boldsymbol{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\|)=\Phi\left(-z_\alpha+\frac{n\kappa(1-\kappa)\|\boldsymbol{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\|^2}{\sqrt{2\mathrm{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Lambda})^2}}\right),$$

where $\tilde{\boldsymbol{\Sigma}}=(1-\kappa)\boldsymbol{\Sigma}_1+\kappa\boldsymbol{\Sigma}_2$. In contrast, Chen and Qin (2010) showed that the power of their proposed test is

$$\beta_{CQ}(\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|)=\Phi\left(-z_\alpha+\frac{n\kappa(1-\kappa)\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2}{\sqrt{2\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^2)}}\right).$$

Thus, the asymptotic relative efficiency of our test with respect to CQ is

$$\mathrm{ARE(FS,CQ)}=\frac{\|\boldsymbol{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\|^2}{\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2}\sqrt{\frac{\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^2)}{\mathrm{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Lambda})^2}}.$$

Here we consider some representative cases.

(i) $\mu_{1k} - \mu_{2k} = \delta,\ k = 1, \dots, p$. Then,

$$\mathrm{ARE(FS, CQ)} = \frac{\mathrm{tr}(\Lambda^2)}{p}\sqrt{\frac{\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^2)}{\mathrm{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Lambda})^2}} \geq 1$$

by the Cauchy inequality. When the variances of the components are equal, the tests are equivalently powerful from the asymptotic viewpoint. Otherwise, the proposed test is preferable in terms of asymptotic power under local alternatives.

(ii) $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, diagonal. The variances of two half of components are ζ_1^2 and ζ_2^2. Assume $\mu_{1k} - \mu_{2k} = \delta$ and $k = 1, \dots, \lfloor \frac{p}{2} \rfloor$. Then,

$$\mathrm{ARE(FS, CQ)} = \frac{\sqrt{\zeta_1^4 + \zeta_2^4}}{\sqrt{2}\zeta_1^2}.$$

Therefore, the proposed test is more powerful than CQ if $\zeta_1^2 < \zeta_2^2$, and vice versa. The ARE has a positive lower bound of $1/\sqrt{2}$ when $\zeta_1^2 \gg \zeta_2^2$. It can be arbitrarily large if ζ_1^2/ζ_2^2 is close to zero, showing the need for the scalar-invariance test.

3. Simulation

Here we report a simulation study designed to evaluate the performance of our proposed test (abbreviated as FS). We compare our tests with the method proposed by Chen and Qin (2010) (abbreviated as CQ), and Srivastava et al. (2013) (abbreviated as SKK), and Park and Ayyala (2013) (abbreviated as PA) under the unequal covariance matrices assumption. We consider the following moving average model as Chen and Qin (2010):

$$X_{ijk} = \rho_{i1}Z_{ij} + \rho_{i2}Z_{i(j+1)} + \cdots + \rho_{iL_i}Z_{i(j+L_i-1)} + \mu_{ij}$$

for $i = 1, 2,\ j = 1, \dots, n_i$ and $k = 1, \dots, p$, where $\{Z_{ijk}\}$ are, respectively, i.i.d. random variables. Consider two scenarios for the innovation $\{Z_{ijk}\}$: (Scenario I) all the $\{Z_{ijk}\}$ are from $N(0, 1)$; (Scenario II) the first half of the components of $\{Z_{ijk}\}_{k=1}^p$ are from centralized Gamma(4,1) so that it has zero mean, and the second half of the components are from $N(0, 1)$. The coefficients $\{\rho_{il}\}_{l=1}^{L_i}$ are generated independently from $U(2, 3)$ and are kept fixed once they are generated through our simulations. The correlations among X_{ijk} and X_{ijl} are determined by $|k - l|$ and L_i. We choose $L_1 = 3$, and $L_2 = 4$ to generate the different covariances of $\mathbf{X}_i$.

We examine the empirical sizes and the estimation efficiency of the tests. Under the null hypothesis, the components of the common vector $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0 = (\mu_1, \dots, \mu_p)$ are generated from $U(0, \lambda)$. The sample sizes are $n_1 = n_2 = 15$. First, we consider the impact of the dimensions. We fix $\lambda = 10$ and consider the six dimensions of $p = 25, 50, 100, 200, 400$ and 800. We summarize the simulation results using the mean-standard deviation-ratio (MDR) $E(T)/\sqrt{\mathrm{var}(T)}$ and the variance ratio (VR) $\widehat{\mathrm{var}(T)}/\mathrm{var}(T)$. Because the explicit forms of $E(T)$ and $\mathrm{var}(T)$ are difficult to calculate, we estimate them by simulation. Fig. 1 shows the MDR, VR and empirical sizes of these four tests with different dimensions. We observe that the MDR and VR of the SKK test are larger than zero and one when the dimensions becomes larger. This is not strange because SKK requires that the dimensions are a smaller order of n^2. Second, we consider the impact of common shifts. We fix the dimension $p = 800$ and consider five common shifts $\lambda = 10, 20, 30, 40$ and 50. Fig. 2 reports the MDR, VR and empirical sizes of these four tests with different common shifts. The MDR and VR of the PA test become larger when the common shifts are larger, which further demonstrates that the PA test is not shift-invariant. In contrast, the MDR and VR of our test are approximately zero and one, respectively. We can control the empirical size well. However, the empirical sizes of the other three tests deviate from the nominal level in many cases.

Next, we compare the power of all the tests. Here, we only report the case of $n_1 = n_2 = 15$ and $p = 800$. For the alternative hypothesis, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\mu}$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is generated as above. We choose $\boldsymbol{\mu}$ in two scenarios: Case A allocates all of the components of equal magnitude to be nonzero; Case B randomly allocates half of the components of equal magnitude to be nonzero. To make the power comparable among the configurations of H_1, we set $\eta := \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \mathrm{tr}(\boldsymbol{\Sigma}_2^2)} = 0.15, 0.2, 0.25, 0.3, 0.35$ throughout the simulation. Fig. 3 reports the empirical power of these four tests. Under Scenario II, CQ is less powerful than the other three tests which is consistent with the theoretical results in Section 2. Furthermore, our test performs better than the SKK and PA tests in all cases. These results suggest that the newly proposed FS test is efficient for testing the equality of locations and is particularly useful when the variances of components are not equal and the dimensions are ultra-high.

Finally, we compare our test with two other tests, the test proposed by Ahmad (2014) (abbreviated as AH) and the test proposed by Thulin (2014) (abbreviated as RS). Here, we consider a multivariate normal distribution with equal covariance. i.e. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Two scenarios are considered for $\boldsymbol{\Sigma}$: (i) weak dependence: $\boldsymbol{\Sigma} = (0.5^{|i-j|})$ and (ii) strong dependence: $\boldsymbol{\Sigma} = (a_{ij}),\ a_{ii} = 1,\ a_{ij} = 0.5,\ i \neq j$. The sample sizes are not equal, i.e. $n_1 = 15$ and $n_2 = 20$. We consider two dimensions $p = 50$ and 100. For the alternative hypothesis, the settings are the same as above, except $\boldsymbol{\mu}_0 \sim U(0, 2)$. Table 1 reports the results of the empirical sizes and power with $\eta = 0.15$ and 0.3 under Case A. The empirical sizes of AH are smaller than the nominal level because AH is not a shift-invariant test. Under the null hypothesis, the expectation of the AH test statistic is $(n_1 + n_2 - 2\sqrt{n_1 n_2} - 2)\boldsymbol{\mu}_0^T\boldsymbol{\mu}_0$. In this case, $n_1 + n_2 - 2\sqrt{n_1 n_2} - 2$ is negative. Consequently, the power of AH is smaller than

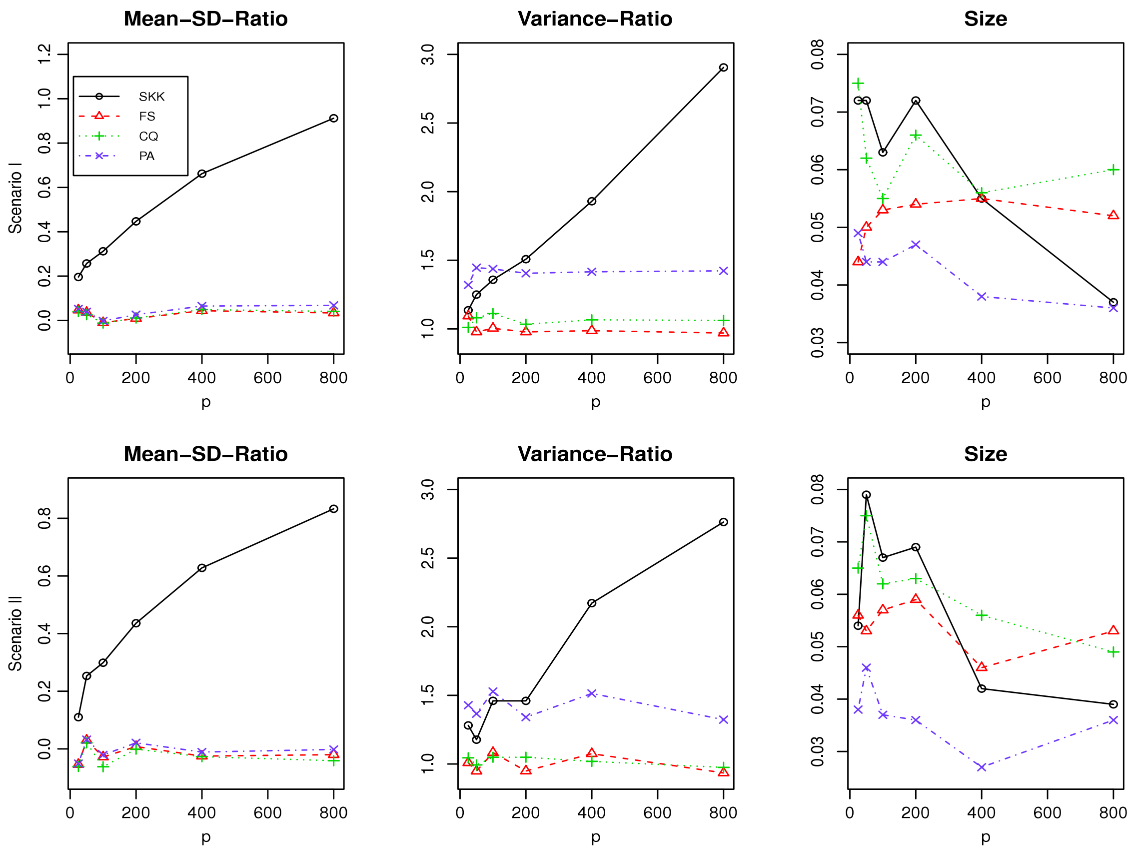

Fig. 1. The MDR, VR and empirical sizes of tests with different dimensions.

Table 1
Empirical sizes and power (%) comparison at 0.05 significance.

p	η	Weak dependence			Strong dependence		
		AH	FS	RS	AH	FS	RS
50	0.00	0.0	6.1	4.9	2.7	6.6	4.3
	0.15	7.7	31.6	20.6	30.2	30.8	13.2
	0.30	28.7	60.5	42.2	48.0	48.3	21.1
100	0.00	0.0	4.2	4.4	3.4	6.1	4.1
	0.15	1.9	31.6	25.3	27.0	27.8	16.9
	0.30	13.4	64.0	50.5	47.9	48.3	29.0

our FS test. The empirical sizes of our FS test are a little larger than the nominal level under the strong dependence case. The power of our FS test is eventually larger than that of the RS test in all cases. By only choosing random subspaces, the RS test may not perform well in these cases.

4. Discussion

Our asymptotic and numerical results suggest that the proposed test is efficient in testing the equality of locations. However, we lose all the information of the correlation between those variables. How to construct a more powerful test under the strong dependence case deserves further research. Furthermore, our test is based on L_2-norm, which is powerful for denser and fainter signals but not efficient for sparser and stronger signals. In a significant development in another direction using the max-norm rather than the L_2-norm, Cai et al. (2013) proposed a test based on the max-norm of marginal t-statistics. See also Zhong et al. (2013) for a related discussion.

Acknowledgments

This work was supported by the National Natural Science Foundation of China grants 11101074 and 11471069 and the Fundamental Research Funds for the Central Universities grant 242015KJ003.

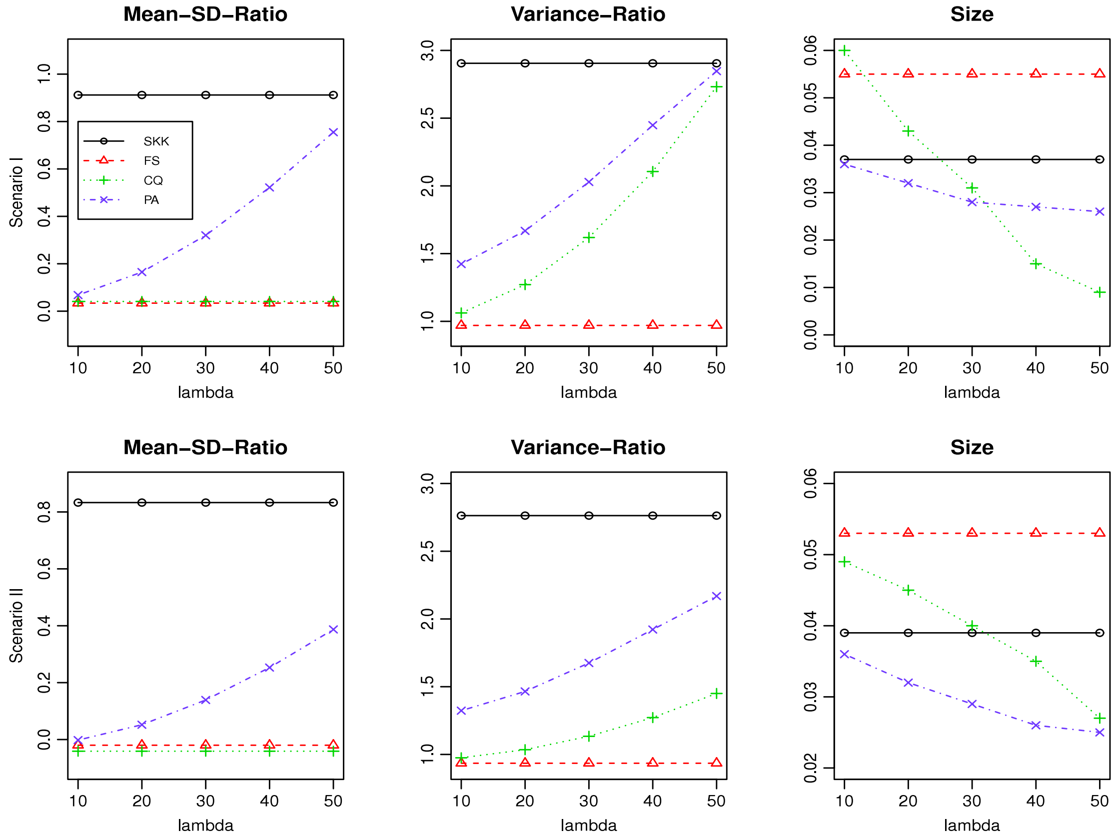

Fig. 2. The MDR, VR and empirical sizes of tests with different λ.

Appendix. Proof of Theorem 1

We decompose T_n into two parts,

$$\begin{aligned} T_n &= \frac{1}{n_1(n_1-1)}\frac{1}{n_2(n_2-1)}\sum_{k=1}^{p}\sum^{n_1}\sum^{n_1}_{i\neq j}\sum^{n_2}\sum^{n_2}_{s\neq t}\frac{(X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk})}{\sigma_{1k}^2+\gamma\sigma_{2k}^2} \\ &\quad+\frac{1}{n_1(n_1-1)}\frac{1}{n_2(n_2-1)}\sum_{k=1}^{p}\sum^{n_1}\sum^{n_1}_{i\neq j}\sum^{n_2}\sum^{n_2}_{s\neq t}(X_{1ik}-X_{2sk})(X_{1jk}-X_{2tk})\left(\frac{1}{\hat\sigma_{1k(i,j)}^2+\gamma\hat\sigma_{2k(s,t)}^2}-\frac{1}{\sigma_{1k}^2+\gamma\sigma_{2k}^2}\right) \\ &\doteq T_{n1}+T_{n2}. \end{aligned}$$

It is straightforward to see that

$$E(T_{n1})=\sum_{k=1}^{p}\frac{(\mu_{1k}-\mu_{2k})^2}{\sigma_{1k}^2+\gamma\sigma_{2k}^2}=\|\boldsymbol{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\|^2,$$

$$\begin{aligned} \mathrm{var}(T_{n1}) &= \frac{2}{n_1(n_1-1)}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda})^2)+\frac{2}{n_2(n_2-1)}\mathrm{tr}((\boldsymbol{\Lambda}\boldsymbol{\Sigma}_2\boldsymbol{\Lambda})^2)+\frac{4}{n_1n_2}\mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_2\boldsymbol{\Lambda}) \\ &\quad+\frac{4}{n_1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_1\boldsymbol{\Lambda}^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)+\frac{4}{n_2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Lambda}^2\boldsymbol{\Sigma}_2\boldsymbol{\Lambda}^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2). \end{aligned}$$

Lemma 1. *Under the same conditions as Theorem 1, as p and $n\to\infty$,*

$$\frac{T_{n1}-E(T_{n1})}{\sqrt{\mathrm{var}(T_{n1})}}\xrightarrow{\mathcal{L}}N(0,1).$$

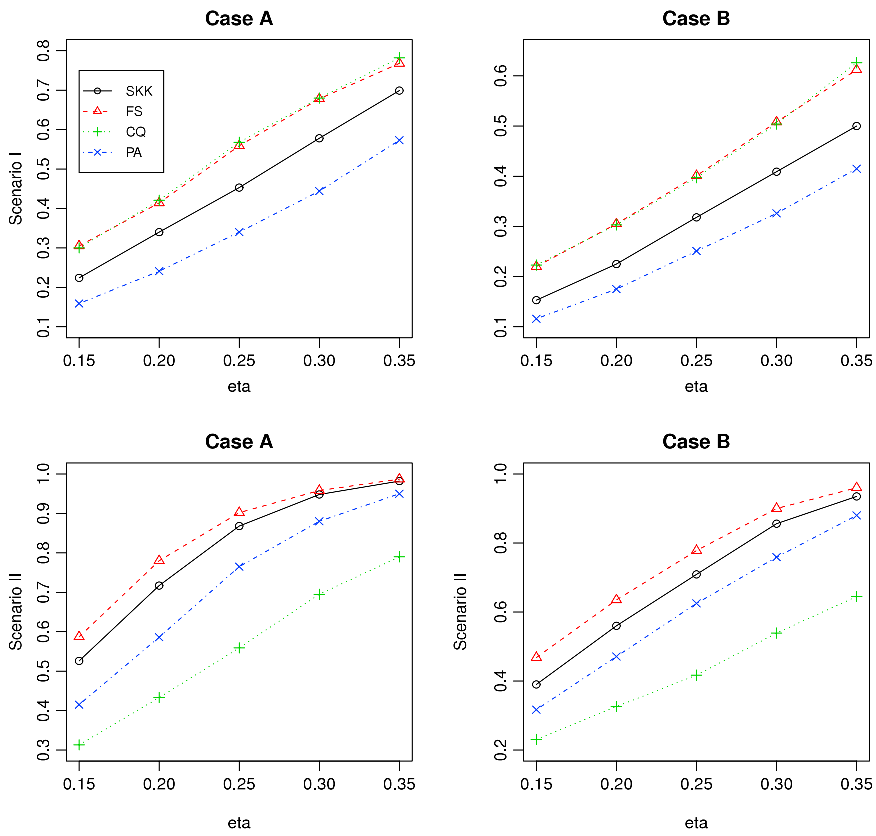

Fig. 3. The power of tests with different η when $n_1 = n_2 = 15$, $p = 800$.

This lemma is a direct corollary of Theorem 1 in Chen and Qin (2010).

Next, we only need to show that $T_{n2} = o_p(\sqrt{\mathrm{var}(T_{n1})})$. Define $Y_{ijk} = X_{ijk} - \mu_{ik},\ i = 1, 2, j = 1, \ldots, n_i, k = 1, \ldots, p$.

$$
\begin{aligned}
T_{n2} = {} & \frac{1}{n_1(n_1-1)}\frac{1}{n_2(n_2-1)}\sum_{k=1}^{p}\sum_{i\neq j}^{n_1}\sum^{n_1}\sum_{s\neq t}^{n_2}\sum^{n_2}(Y_{1ik}-Y_{2sk})(Y_{1jk}-Y_{2tk})\left(\frac{1}{\hat{\sigma}^2_{1k(i,j)}+\gamma\hat{\sigma}^2_{2k(s,t)}}-\frac{1}{\sigma^2_{1k}+\gamma\sigma^2_{2k}}\right)\\
& + \frac{2}{n_1n_2}\sum_{k=1}^{p}\sum_{i=1}^{n_1}\sum_{s=1}^{n_2}(Y_{1ik}-Y_{2sk})(\mu_{1k}-\mu_{2k})\left(\frac{1}{\hat{\sigma}^2_{1k(i,j)}+\gamma\hat{\sigma}^2_{2k(s,t)}}-\frac{1}{\sigma^2_{1k}+\gamma\sigma^2_{2k}}\right)\\
& + \sum_{k=1}^{p}(\mu_{1k}-\mu_{2k})^2\left(\frac{1}{\hat{\sigma}^2_{1k(i,j)}+\gamma\hat{\sigma}^2_{2k(s,t)}}-\frac{1}{\sigma^2_{1k}+\gamma\sigma^2_{2k}}\right)\\
\doteq {} & R_1 + R_2 + R_3.
\end{aligned}
$$

Define $\mathbf{\Lambda}_{(i,j,s,t)} = \mathrm{diag}\{(\hat{\sigma}^2_{11(i,j)}+\gamma\hat{\sigma}^2_{21(s,t)})^{-1/2}, \ldots, (\hat{\sigma}^2_{1p(i,j)}+\gamma\hat{\sigma}^2_{2p(s,t)})^{-1/2}\}$.

$$
\begin{aligned}
R_1 = {} & \frac{1}{n_2(n_2-1)}\sum_{t\neq s}^{n_2}\sum_{s=1}^{n_2}\left(\frac{1}{n_1(n_1-1)}\sum_{j\neq i}^{n_1}\sum_{i=1}^{n_1}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}^2_{(i,j,s,t)}-\mathbf{\Lambda}^2)\mathbf{Y}_{1j}\right)\\
& + \frac{1}{n_1(n_1-1)}\sum_{j\neq i}^{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{n_2(n_2-1)}\sum_{t\neq s}^{n_2}\sum_{s=1}^{n_2}\mathbf{Y}_{2s}^T(\hat{\mathbf{\Lambda}}^2_{(i,j,s,t)}-\mathbf{\Lambda}^2)\mathbf{Y}_{2t}\right)\\
& - \frac{2}{(n_1-1)(n_2-1)}\sum_{j\neq i}^{n_1}\sum_{s\neq t}^{n_2}\left(\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{s=1}^{n_2}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}^2_{(i,j,s,t)}-\mathbf{\Lambda}^2)\mathbf{Y}_{2t}\right).
\end{aligned}
$$

By Theorem 1 in Park and Ayyala (2013), we have

$$
E\left(\frac{1}{n_1(n_1-1)}\sum_{j\neq i}^{n_1}\sum_{i=1}^{n_1}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}^2_{(i,j,s,t)}-\mathbf{\Lambda}^2)\mathbf{Y}_{1j}\right)^2 = O(n^{-3}\mathrm{tr}((\mathbf{\Lambda}\mathbf{\Sigma}_1\mathbf{\Lambda})^2)) = o(\mathrm{var}(T_{n1}))
$$

$$E\left(\frac{1}{n_2(n_2-1)}\sum_{t\neq s}^{n_2}\sum_{s=1}^{n_2}\mathbf{Y}_{2s}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)\mathbf{Y}_{2t}\right)^2=O(n^{-3}\mathrm{tr}((\mathbf{\Lambda}\mathbf{\Sigma}_2\mathbf{\Lambda})^2))=o(\mathrm{var}(T_{n1}))$$
$$E\left(\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{s=1}^{n_2}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)\mathbf{Y}_{2t}\right)^2=O(n^{-3}\mathrm{tr}(\mathbf{\Lambda}\mathbf{\Sigma}_1\mathbf{\Lambda}^2\mathbf{\Sigma}_2\mathbf{\Lambda}))=o(\mathrm{var}(T_{n1})).$$

Thus, $R_1=o_p(\sqrt{\mathrm{var}(T_{n1})})$. Similarly,

$$R_2=\frac{2}{n_2(n_2-1)(n_1-1)}\sum_{t\neq s}^{n_2}\sum_{s=1}^{n_2}\sum_{j\neq i}^{n_1}\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)$$
$$-\frac{2}{n_1(n_1-1)(n_2-1)}\sum_{j\neq i}^{n_1}\sum_{i=1}^{n_1}\sum_{t\neq s}^{n_2}\left(\frac{1}{n_2}\sum_{s=1}^{n_1}\mathbf{Y}_{2s}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right).$$

By the proof of Theorem 2 in Park and Ayyala (2013), we have

$$E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{Y}_{1i}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)^2=O(n^{-2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\mathbf{\Lambda}^2\mathbf{\Sigma}_1\mathbf{\Lambda}^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2))$$
$$E\left(\frac{1}{n_2}\sum_{s=1}^{n_1}\mathbf{Y}_{2s}^T(\hat{\mathbf{\Lambda}}_{(i,j,s,t)}^2-\mathbf{\Lambda}^2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)^2=O(n^{-2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\mathbf{\Lambda}^2\mathbf{\Sigma}_2\mathbf{\Lambda}^2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)).$$

By the Condition (C3), we also have $R_2=o_p(\sqrt{\mathrm{var}(T_{n1})})$. And $E(R_3^2)=O(n^{-1}((\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\mathbf{\Lambda}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2))^2)=o(\mathrm{var}(T_{n1}))$. Thus, we proof that $T_{n2}=o_p(\sqrt{\mathrm{var}(T_{n1})})$.

References

Ahmad, R.M., 2014. A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens–Fisher setting. Ann. Inst. Statist. Math. 66, 33–61.
Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statist. Sinica 6, 311–329.
Cai, T., Liu, W.D., Xia, Y., 2013. Two-sample test of high dimensional means under dependency. J. R. Stat. Soc. Ser. B 76, 349–372.
Chang, J., Zhou, W., Zhou, W.X., 2014. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. arXiv:1406.1939.
Chen, S.X., Qin, Y.-L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist. 38, 808–835.
Feng, L., Zou, C., Wang, Z., Zhu, L., 2015. Two sample Behrens–Fisher problem for high-dimensional data. Statist. Sinica in press.
Gregory, K.B., Carroll, R.J., Baladandayuthapani, V., Lahiri, S.N., 2015. A two-sample test for equality of means in high dimension. J. Amer. Statist. Assoc. in press.
Ley, C., Paindaveine, D., Verdebout, T., 2015. High-dimensional tests for spherical location and spiked covariance. J. Multivariate Anal. 139, 79–91.
Li, J., Chen, S.X., 2012. Two sample tests for high-dimensional covariance matrices. Ann. Statist. 40, 908–940.
Nettleton, D., Recknor, J., Reecy, J., 2008. Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis. Bioinformatics 24, 192–201.
Paindaveine, D., Verdebout, T., 2013. Universal asymptotics for high-dimensional sign tests, arXiv:1311.3118v1.
Park, J., Ayyala, D.N., 2013. A test for the mean vector in large dimension and small samples. J. Statist. Plann. Inference 143, 929–943.
Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. J. Multivariate Anal. 99, 386–402.
Srivastava, M.S., Katayama, S., Kano, Y., 2013. A two sample test in high dimensional data. J. Multivariate Anal. 114, 349–358.
Thulin, M., 2014. A high-dimensional two-sample test for the mean using random subspaces. Comput. Statist. Data Anal. 74, 26–38.
Zhong, P.-S., Chen, S.X., Xu, M., 2013. Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. Ann. Statist. 41, 2820–2851.