



Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension

Shota Katayama^{a,*}, Yutaka Kano^a, Muni S. Srivastava^b

^a Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

^b Department of Statistics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

ARTICLE INFO

Article history:

Received 16 December 2010

Available online 25 January 2013

AMS subject classifications:

primary 62E20

secondary 62H15

Keywords:

Hypothesis testing

High-dimensional data

Multivariate normal distribution

Asymptotic theory

ABSTRACT

The problem of hypothesis testing concerning the mean vector for high dimensional data has been investigated by many authors. They have proposed several test criteria and obtained their asymptotic distributions, under somewhat restrictive conditions, when both the sample size and the dimension tend to infinity. Indeed, the conditions used by these authors exclude a typical situation where the population covariance matrix has spiked eigenvalues, as for instance, the population covariance matrix with the compound symmetry structure (the variances are the same; the covariances are the same). In this paper, we relax their conditions to include such important cases, obtaining rather non-standard asymptotic distributions which are the convolution of normal and chi-squared distributions for the population covariance matrix with moderate spiked eigenvalues, and obtaining the asymptotic distributions in the form of convolutions of chi-square distributions for the population covariance matrix with quite spiked eigenvalues.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

We consider the problem of hypothesis testing concerning the mean vector under normality. Suppose that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independently and identically distributed (i.i.d.) random vectors drawn from a p -dimensional multivariate normal distribution $N_p(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$, where $\boldsymbol{\Sigma}_p$ is positive definite. The sample mean vector $\bar{\mathbf{X}}_n$ and the sample covariance matrix \mathbf{S}_n are typically defined as

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad \mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)^T.$$

Consider the following test of hypothesis:

$$H_0 : \boldsymbol{\mu}_p = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_p \neq \mathbf{0}. \quad (1.1)$$

When $n > p$, typically one applies Hotelling's T^2 to test the above hypothesis, where T^2 is defined as

$$T^2 = n \bar{\mathbf{X}}_n^T \mathbf{S}_n^{-1} \bar{\mathbf{X}}_n.$$

However, this statistic cannot be defined for the case where the dimension of data is larger than or equal to the sample size, i.e., $n \leq p$, because \mathbf{S}_n^{-1} does not exist. For this case, Dempster [9] proposes an alternative test statistic where $\text{tr} \mathbf{S}_n$

* Corresponding author.

E-mail address: sfujimoto@sigmath.es.osaka-u.ac.jp (S. Katayama).

replaces \mathbf{S}_n :

$$T_D^2 = \frac{n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n}{\text{tr } \mathbf{S}_n}.$$

The distribution of the test statistic T_D^2 is not similar to that of T^2 . It depends on Σ_p which appears as a nuisance parameter. Dempster [9,10], Bai and Saranadasa [2], Fujikoshi et al. [12] and Srivastava [24] have developed approximate and asymptotic distributions of T_D^2 .

Under the null hypothesis H_0 , Dempster [9,10] shows that the statistic T_D^2 is approximately distributed as $F_{r,(n-1)r}$, with

$$r = \frac{(\text{tr } \Sigma_p)^2}{\text{tr } \Sigma_p^2}, \quad (1.2)$$

where $F_{a,b}$ denotes a F -distribution with a and b degrees of freedom. In general, r is unknown since it is a function of Σ_p which is usually unknown. Dempster [9] provides two estimators of r . A problem of these estimators is that they are not invariant under orthogonal transformation of the data matrix. Srivastava [24] proposes a different estimator for r which avoids this problem. The estimator will be described in Section 2.

When both the sample size and the dimension tend to infinity, Bai and Saranadasa [2] derive an asymptotic normal distribution for T_D^2 under the null hypothesis H_0 and the condition

$$\max_{1 \leq i \leq p} \lambda_i = o\left(\sqrt{\text{tr } \Sigma_p^2}\right), \quad (1.3)$$

where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of Σ_p . It may also be mentioned that Bai and Saranadasa [2] propose an alternative version to T_D^2 defined by

$$T_{BS}^2 = n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n - \text{tr } \mathbf{S}_n.$$

The asymptotic normality of T_{BS}^2 has also been obtained under H_0 and the condition in (1.3). More recently, Fujikoshi et al. [12] and Srivastava [24] have obtained the asymptotic null distributions of T_D^2 and T_{BS}^2 under the different condition that

$$0 < \lim_{p \rightarrow \infty} \frac{\text{tr } \Sigma_p^i}{p} < \infty, \quad i = 1, 2, 3, 4. \quad (1.4)$$

It is noted that T_D^2 and T_{BS}^2 are invariant under the transformation $\mathbf{X}_i \rightarrow c\Gamma\mathbf{X}_i$ where $c > 0$ and $\Gamma^T\Gamma = \mathbf{I}_p$, where \mathbf{I}_p is the $p \times p$ identity matrix. It should be noted that these criteria are not invariant under the scale transformation $\mathbf{X}_i \rightarrow \mathbf{D}\mathbf{X}_i$ where $\mathbf{D} = \text{diag}(d_1, \dots, d_p)$ with $d_i > 0$.

Srivastava and Du [25] propose an invariant test criterion under the scale transformation given by

$$T_S^2 = n\bar{\mathbf{X}}_n^T \mathbf{D}_S^{-1} \bar{\mathbf{X}}_n,$$

where

$$\mathbf{D}_S = \text{diag}(s_{11}, \dots, s_{pp}), \quad \mathbf{S}_n(i, j) = s_{ij}.$$

When both n and p tend to infinity, they establish its asymptotic normality under H_0 and the condition that

$$0 < \lim_{p \rightarrow \infty} \frac{\text{tr } \mathcal{R}_p^i}{p} < \infty, \quad i = 1, 2, 3, 4, \quad (1.5)$$

where \mathcal{R}_p is the population correlation matrix.

Either of the conditions in (1.3) and (1.4) implies $r \rightarrow \infty$ as $p \rightarrow \infty$. It follows from the Cauchy–Schwarz inequality that $r \leq p$, with equality if and only if $\Sigma_p = c\mathbf{I}_p$ for some constant $c > 0$. From these findings, we would say that these conditions require Σ_p to be almost proportional to the identity matrix. We can say the same thing for the population correlation matrix \mathcal{R}_p satisfying the condition in (1.5).

For instance, the conditions in (1.3)–(1.5) are satisfied for each of the following two cases:

- (a) Sphericity model: $\Sigma_p = c\mathbf{I}_p$ where $c > 0$.
- (b) Autoregressive model: $\Sigma_p(i, j) = c(\rho^{|i-j|})$ where $c > 0$ and $0 < \rho < 1$.

We apply Theorem 9 in Gray [13] to show that the case (b) meets the conditions in (1.3)–(1.5). This theorem states that if the $p \times p$ Toeplitz matrices

$$\mathbf{T}_p = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(p-1)} \\ t_1 & t_0 & t_{-1} & \cdots & t_{-(p-2)} \\ t_2 & t_1 & t_0 & \cdots & t_{-(p-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{p-1} & t_{p-2} & t_{p-3} & \cdots & t_0 \end{pmatrix}$$

satisfy $\sum_{k=-\infty}^{\infty} |t_k| < \infty$, it holds that

$$\lim_{p \rightarrow \infty} \frac{\text{tr} \mathbf{T}_p^k}{p} = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda)^k d\lambda, \quad \text{where } f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda}. \quad (1.6)$$

On the other hand, the conditions in (1.3)–(1.5) are not satisfied for the following cases:

(c) Compound symmetry structure: $\Sigma_p = c(1 - \rho)\mathbf{I}_p + c\rho\mathbf{1}_p\mathbf{1}_p^T$ where $c > 0$, $0 < \rho < 1$ and $\mathbf{1}_p$ is the p column vector with all entries one.

Note that the eigenvalues of the compound symmetry structure are $\lambda_1 = c(1 - \rho + p\rho)$, $\lambda_2 = \dots = \lambda_p = c(1 - \rho)$. This structure is a special case of

(d) Spiked model: $\lambda_i = c_i p^{\alpha_i}$ ($i = 1, \dots, \ell$), $\lambda_j = c_j$ ($j = \ell + 1, \dots, p$), where $\lambda_1 \geq \dots \geq \lambda_p$ are eigenvalues of Σ_p (or \mathcal{R}_p) and $c_i(>0)$, $\alpha_i(>0)$, $m(<p)$ are all independent of p .

Suppose that $c_i = 1$ ($i = 1, \dots, p$). The condition in (1.3) is satisfied when $\alpha_1 < 1/2$, the condition in (1.4) is satisfied when $\alpha_1 \leq 1/4$, and neither of them is true when $\alpha_1 \geq 1/2$, which implies that these conditions are satisfied only when the divergence speeds of the largest eigenvalues of Σ_p (or \mathcal{R}_p) are very slow.

The following new condition concerning Σ_p (or \mathcal{R}_p) is introduced which includes the above cases (a)–(d); there exist positive numbers $\delta_i > 0$ ($i = 2, 3, \dots$) with $\delta_2 \geq \delta_k$ ($k = 3, 4, \dots$) such that

$$0 < \lim_{p \rightarrow \infty} \text{tr} \left(\frac{\Sigma_p}{p^{\delta_i}} \right)^i < \infty, \quad i = 2, 3, \dots \quad (1.7)$$

Obviously, the cases (a) and (b) meet the condition in (1.7) with $\delta_i = 1/i$, the case (c) meets the condition with $\delta_i = 1$, and it shall be seen in Section 2 that the case (d) also meets the condition. As seen above, there are some covariance matrices or its correlation matrices which do not meet the conditions in (1.3), (1.4) nor (1.5) but meet the condition in (1.7). Strictly speaking, however, the condition in (1.7) is not mathematically weaker than the conditions in (1.3)–(1.5) since these conditions make no restrictions on the i th power of Σ_p (or \mathcal{R}_p) for $i > 4$. We suppose that there are not so many important cases which do not meet (1.7) but do (1.3), (1.4) or (1.5).

The main purpose of this paper is to develop the asymptotic distributions for the statistics T_D^2 , T_{BS}^2 and T_S^2 under the condition in (1.7) when both the sample size and the dimension tend to infinity. This is done in Sections 2 and 3. In Section 4, we numerically compare these asymptotic distributions with an already existing asymptotic distribution and show that the new asymptotic distributions exhibit better type one errors.

2. Asymptotic distributions for quadratic form

We first derive the asymptotic distribution for a certain quadratic form when the dimension tends to infinity. As we shall see later in Section 3, under H_0 in (1.1), each of the three test criteria T_D^2 , T_{BS}^2 and T_S^2 , proposed by Dempster [9], Bai and Saranadasa [2] and Srivastava and Du [25] respectively, is asymptotically equivalent to the quadratic form.

Theorem 2.1. Let \mathbf{z}_p be a random vector distributed as $N_p(\mathbf{0}, \Sigma_p)$ where Σ_p is positive definite and satisfies the condition in (1.7). Let $\phi(t)$ be the characteristic function of

$$Z_p = \frac{\mathbf{z}_p^T \mathbf{z}_p - \text{tr} \Sigma_p}{\sqrt{2 \text{tr} \Sigma_p^2}}. \quad (2.1)$$

Then for all $|t| < 1/2$, it holds that

$$\log \phi(t) \rightarrow \frac{(it)^2}{2} + \sum_{k \in \Delta} \frac{1}{k} \frac{2^{k-1} (it)^k m_k}{(2m_2)^{k/2}}, \quad p \rightarrow \infty,$$

where

$$m_i = \lim_{p \rightarrow \infty} \text{tr} \left(\frac{\Sigma_p}{p^{\delta_i}} \right)^i, \quad \Delta = \{k \mid \delta_2 = \delta_k, k \geq 3\}. \quad (2.2)$$

Proof. Let $\mathbf{u}_p = \Sigma_p^{-1/2} \mathbf{z}_p$; then \mathbf{u}_p is distributed as $N(\mathbf{0}, \mathbf{I}_p)$. Let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of Σ_p and let U_1, \dots, U_p be i.i.d. random variables with chi-square distribution with 1 degree of freedom. Since $\mathbf{z}_p^T \mathbf{z}_p = \sum_{j=1}^p \lambda_j U_j$, we obtain

$$\log \phi(t) = -\frac{1}{2} \sum_{j=1}^p \log \left(1 - \frac{2it\lambda_j}{\sqrt{2 \text{tr} \Sigma_p^2}} \right) - \frac{it \text{tr} \Sigma_p}{\sqrt{2 \text{tr} \Sigma_p^2}}.$$

Generally, for any complex number z with $|z| < 1$, it holds that

$$-\log(1-z) = \sum_{k=1}^{\infty} \frac{z^k}{k}.$$

Using the above, when $|t| < 1/2$, $\log \phi(t)$ is expanded as

$$\log \phi(t) = \sum_{k=1}^{\infty} \frac{1}{k} \frac{2^{k-1}(it)^k \text{tr} \Sigma_p^k}{(2\text{tr} \Sigma_p^2)^{k/2}} - \frac{it \text{tr} \Sigma_p}{\sqrt{2\text{tr} \Sigma_p^2}} = \sum_{k=2}^{\infty} \frac{1}{k} \frac{2^{k-1}(it)^k \text{tr} \Sigma_p^k}{(2\text{tr} \Sigma_p^2)^{k/2}}.$$

When $|t| < 1/2$, this infinite series converges uniformly in $p \geq 0$, since we have

$$\left| \frac{1}{k} \frac{2^{k-1}(it)^k \text{tr} \Sigma_p^k}{(2\text{tr} \Sigma_p^2)^{k/2}} \right| \leq \frac{2^{k-1}|t|^k}{k},$$

which is independent of p and it also holds that

$$\sum_{k=1}^{\infty} \frac{2^{k-1}|t|^k}{k} = -\frac{1}{2} \log(1-2|t|).$$

Hence we obtain

$$\lim_{p \rightarrow \infty} \log \phi(t) = \sum_{k=2}^{\infty} \lim_{p \rightarrow \infty} \frac{1}{k} \frac{2^{k-1}(it)^k \text{tr} \Sigma_p^k}{(2\text{tr} \Sigma_p^2)^{k/2}} = \frac{(it)^2}{2} + \sum_{k \in \Delta} \frac{1}{k} \frac{2^{k-1}(it)^k m_k}{(2m_2)^{k/2}}$$

in view of (2.2). The proof of the theorem is completed. \square

From the characteristic functions of the standard normal distribution and chi-square distribution, we obtain the following corollary.

Corollary 2.1. *If the set Δ in (2.2) is empty, then the asymptotic distribution for Z_p is the standard normal distribution $N(0, 1)$. If $\Delta = \{3, 4, \dots\}$ and $m_i = c^i m$ ($i = 2, 3, \dots$) for some positive constants c and m , then the asymptotic distribution is the standardized chi-square distribution with m degrees of freedom, that is, $(\chi_m^2 - m)/\sqrt{2m}$.*

2.1. Asymptotic distributions for several population covariance matrices

From Theorem 2.1 and Corollary 2.1, it is noted that the type of the asymptotic distribution for Z_p in (2.1) depends on the population covariance matrix Σ_p . For this reason, we define the asymptotic distribution for Z_p as $\Phi(\Sigma_p)$ and define the convergence in distribution as

$$Z_p \xrightarrow{d} \Phi(\Sigma_p). \quad (2.3)$$

We shall show some examples illustrating $\Phi(\Sigma_p)$ for different types of population covariance matrices.

Example 2.1 (Sphericity model and autoregressive model). Let Σ_p be the sphericity model $\Sigma_p = cI_p$ where $c > 0$. Then we have $\delta_i = 1/i$ ($i = 2, 3, \dots$); therefore Δ is empty. It follows from Corollary 2.1 that $\Phi(\Sigma_p) = N(0, 1)$. Similarly, when Σ_p is the autoregressive model $\Sigma_p = c(\rho^{|i-j|})$ ($c > 0$; $0 < \rho < 1$), we have $\Phi(\Sigma_p) = N(0, 1)$ in view of (1.6).

Example 2.2 (Compound symmetry structure). Let Σ_p have the compound symmetry structure $\Sigma_p = c(1-\rho)I_p + c\rho\mathbf{1}_p\mathbf{1}_p^T$ ($c > 0$; $0 < \rho < 1$). Since its eigenvalues are $\lambda_1 = c(1-\rho+\rho p)$, $\lambda_2 = \dots = \lambda_p = c(1-\rho)$, we obtain $\Delta = \{3, 4, \dots\}$ and $m_i = (c\rho)^i$ ($i = 2, 3, \dots$). From Corollary 2.1, we obtain $\Phi(\Sigma_p) = (\chi_1^2 - 1)/\sqrt{2}$. If we now consider the case where Σ_p is the block diagonal matrix such that

$$\Sigma_p = \begin{pmatrix} \Sigma_p^{(1)} & 0 \\ 0 & \Sigma_p^{(2)} \end{pmatrix}$$

where $\Sigma_p^{(i)} = c_i(1-\rho_i)I_{p_i} + c_i\rho_i\mathbf{1}_{p_i}\mathbf{1}_{p_i}^T$ with $c_i > 0$, $0 < \rho_i < 1$ and $p_1 + p_2 = p$, then we note that $\lambda_1 = c_1(1-\rho_1+\rho_1 p_1)$, $\lambda_2 = c_2(1-\rho_2+\rho_2 p_2)$ and that the other eigenvalues are $c_1(1-\rho_1)$ or $c_2(1-\rho_2)$. We assume $p_1/p \rightarrow \eta \in (0, 1]$ as $p \rightarrow \infty$. Then it follows from Theorem 2.1 that

$$\Phi(\Sigma_p) = \frac{c_1\rho_1\eta U_1 - c_1\rho_1\eta}{\sqrt{2(c_1\rho_1\eta)^2 + 2[c_2\rho_2(1-\eta)]^2}} + \frac{c_2\rho_2(1-\eta)U_2 - c_2\rho_2(1-\eta)}{\sqrt{2(c_1\rho_1\eta)^2 + 2[c_2\rho_2(1-\eta)]^2}}, \quad (2.4)$$

where U_1 and U_2 are i.i.d. random variables with chi-square distribution with 1 degree of freedom. The estimators of c_i and ρ_i are given by Donner and Koval [11]. When $c_1 = c_2$, $p_1 = p_2$ and $\rho_1 = \rho_2$, $\Phi(\Sigma_p) = (\chi_2^2 - 2)/\sqrt{2}$.

Example 2.3 (Spiked model). Let the eigenvalues of Σ_p be

$$\lambda_i = c_i p^{\alpha_i} \quad (i = 1, \dots, \ell), \quad \lambda_j = c_j \quad (j = \ell + 1, \dots, p), \quad (2.5)$$

where $c_i (> 0)$, $\alpha_i (> 0)$, $\ell (< p)$ are all independent of p and hold the order $\lambda_1 \geq \dots \geq \lambda_p > 0$. Let $\alpha = \max_{1 \leq i \leq \ell} \alpha_i$ and let q be the number of α_i 's which is equal to α . Suppose that

$$0 < \lim_{p \rightarrow \infty} \frac{\sum_{j=\ell+1}^p c_j^k}{p} < \infty, \quad k = 1, 2, \dots,$$

then the type of the asymptotic distribution for Z_p depends on the following three cases: (i) $0 < \alpha < 1/2$, (ii) $\alpha = 1/2$, (iii) $\alpha > 1/2$.

(i) We can assume $(k+1)^{-1} \leq \alpha < k^{-1}$ with an integer $k (\geq 2)$. Then the condition in (1.7) is satisfied with $\delta_i = 1/i$ ($2 \leq i \leq k$) and $\delta_i = \alpha$ ($i \geq k+1$), which implies that the set Δ in (2.2) is empty. Hence, it follows from Corollary 2.1 that $\Phi(\Sigma_p) = N(0, 1)$.

(ii) Note that the condition in (1.7) is satisfied with $\delta_i = 1/2$ ($i \geq 2$), obtaining that $\Delta = \{3, 4, \dots\}$ and

$$m_2 = \sum_{i=1}^q c_i^2 + d, \quad m_k = \sum_{i=1}^q c_i^k, \quad k = 3, 4, \dots,$$

where $d = \lim_{p \rightarrow \infty} (\sum_{j=\ell+1}^p c_j^2)/p$. From Theorem 2.1, the logarithmic characteristic function of Z_p converges to

$$\frac{(it\sigma)^2}{2} - \frac{1}{2} \sum_{j=1}^q \log \left(1 - \frac{2itc_j}{\sqrt{2m_2}} \right) - \sum_{j=1}^q \frac{itc_j}{\sqrt{2m_2}},$$

which implies

$$Z_p \xrightarrow{d} \Phi(\Sigma_p) = \sigma N + \frac{1}{\sqrt{m_2}} \sum_{j=1}^q c_j \frac{U_j - 1}{\sqrt{2}}, \quad (2.6)$$

where N is a random variable with the standard normal distribution, $\{U_i\}$ are i.i.d. random variables with chi-square distribution with 1 degree of freedom, N and $\{U_i\}$ are mutually independent and

$$\sigma^2 = 1 - \frac{\sum_{j=1}^q c_j^2}{m_2}.$$

When $c_i = c$ ($1 \leq i \leq q$), we have $\Phi(\Sigma_p) = \sigma N(0, 1) + c(\chi_q^2 - q)/\sqrt{2m_2}$. In general, the probability density function of the convolution of normal and chi-squared distributions $N(\mu, \sigma^2) + \gamma \chi_h^2$ is given by (see, e.g., Arunajadai [1])

$$\frac{\sigma^{h/2}}{\sqrt{2\pi\sigma^2(2\gamma)^{h/2}}} \exp \left[-\frac{(y-\mu)^2}{2\sigma^2} + \frac{\sigma^2 z^2}{4} \right] D_{-h/2}(\sigma z), \quad y \in \mathbb{R}^1,$$

where

$$z = \frac{1}{2\gamma} - \frac{(y-\mu)}{\sigma^2},$$

$$D_n(u) = 2^{\frac{n}{2}} e^{\frac{u^2}{4}} \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1-u}{2})} {}_1F_1 \left(-\frac{n}{2}, \frac{1}{2}; \frac{u^2}{2} \right) + \frac{u}{\sqrt{2}} \frac{\Gamma(-\frac{1}{2})}{\Gamma(-\frac{u}{2})} {}_1F_1 \left(\frac{1-u}{2}, \frac{3}{2}; \frac{u^2}{2} \right) \right]$$

where Γ is the Gamma function and ${}_1F_1(a, b; u)$ is the Kummer confluent hypergeometric function given by

$${}_1F_1(a, b; u) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b)}{\Gamma(b+k)\Gamma(a)} \frac{u^k}{k!}.$$

A particular population covariance matrix which originates the convolution of normal and chi-squared distributions is the compound symmetry structure (c) in Section 1 with $\rho = 1/(a + \sqrt{p})$ where a is a constant independent of p . Note that the eigenvalues are given by $\lambda_1 = 1 + \sqrt{p} + o(1)$, $\lambda_2 = \dots = \lambda_p = 1 + o(1)$. In this case, the asymptotic distribution for Z_p is given by $\Phi(\Sigma_p) = N(0, 1) + (\chi_1^2 - 1)/\sqrt{2}$.

(iii) The condition in (1.7) is satisfied with $\delta_i = \alpha$, which implies that $\Delta = \{3, 4, \dots\}$ and $m_k = \sum_{i=1}^q c_i^k$ ($k \geq 2$). From Theorem 2.1, it holds that

$$Z_p \xrightarrow{d} \Phi(\Sigma_p) = \frac{1}{\sqrt{m_2}} \sum_{i=1}^q c_i \frac{U_i - 1}{\sqrt{2}}, \quad (2.7)$$

where U_1, \dots, U_q are i.i.d. random variables with chi-square distribution with 1 degree of freedom. When $c_i = c$, $\Phi(\Sigma_p) = (\chi_q^2 - q)/\sqrt{2q}$.

Remark 2.1. There is a vast literature for finding the distribution functions of the convolutions of chi-square distributions in (2.4) and (2.7), for instance, see Imhof [15] and Davies [7] for the numerical integration, see Shah and Khatri [22] for the power series expansion, and see Kotz et al. [17], Davis [8] and Castaño-Martínez and López-Blásquez [4] for the Laguerre series expansion. The numerical integration methods given by Imhof [15] and Davies [7] also find the distribution function of the convolution of normal and chi-squared distributions in (2.6).

2.2. Rate of convergence

In the previous section, it is shown that the type of the asymptotic distribution for Z_p in (2.1) depends on the structure of Σ_p . Here we shall show that the rate of convergence for Z_p also depends on the structure. For simplicity, we assume that only the first eigenvalue of Σ_p is spiked, that is, Σ_p has the structure (2.5) with $\ell = 1$, $\alpha_1 = \alpha$ and $c_2 = c_j$ ($j = 3, \dots, p$). The result of this section for the general structure (2.5) can be similarly obtained. Define the rate of convergence r_p for Z_p as

$$Z_p = \Phi(\Sigma_p) + O_p(r_p). \quad (2.8)$$

Then we have the following proposition:

Proposition 2.1. Assume that Σ_p has the structure (2.5) with $\ell = 1$, $\alpha_1 = \alpha$ and $c_2 = c_j$ ($j = 3, \dots, p$). Then we have

- (i) $r_p = p^{-1/2}$ when $0 < \alpha \leq 1/2$,
- (ii) $r_p = p^{1/2-\alpha}$ when $\alpha > 1/2$,

where r_p is defined in (2.8).

See Appendix for the proof. Note that the asymptotic distribution for Z_p is the normal distribution or the convolution of normal and chi-squared distributions when $0 < \alpha \leq 1/2$, while it is the chi-square distribution when $\alpha > 1/2$. The proposition shows that the convergence speed of Z_p to the chi-square distribution is faster than that to the other distributions when $\alpha > 1$, is slower when $1/2 < \alpha < 1$, and is equivalent when $\alpha = 1$.

2.3. Approximation for convolutions of chi-square distributions

One may approximate the convolutions of chi-square distributions in (2.4), (2.6) and (2.7) by the methods of Welch [26] and Pearson [19] for instance. For the statistic

$$W = \frac{\sum_{j=1}^q c_j U_j - \sum_{j=1}^q c_j}{\sqrt{2 \sum_{j=1}^q c_j^2}},$$

where c_j ($j = 1, \dots, q$) is a constant and $\{U_j\}$ are i.i.d. as χ_1^2 , Welch [26] and Pearson [19] approximate it as $A_i = (\chi_{h_i}^2 - h_i)/\sqrt{2h_i}$ ($i = 1, 2$) with

$$h_1 = \frac{\left(\sum_{j=1}^q c_j\right)^2}{\sum_{j=1}^q c_j^2}, \quad h_2 = \frac{\left(\sum_{j=1}^q c_j^2\right)^3}{\left(\sum_{j=1}^q c_j^3\right)^2},$$

respectively. See also Satterthwaite [20,21] for the approximation A_1 , and Buckley and Eagleson [3] for the approximation A_2 . In this section, we study how the two approximations work using the measure of proximity given by

$$M_i = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{\phi_W(t) - \phi_{A_i}(t)}{t} \right| dt,$$

Table 1Values of M_i and $M_{i,\alpha}$ for $i = 1, 2$ and $\alpha = 0.95$.

q	(c_1, \dots, c_q)	M_1	M_2	$M_{1,0.95}$	$M_{2,0.95}$
$q = 2$	(0.3, 0.7)	3.389×10^{-2}	1.051×10^{-1}	1.251×10^{-3}	1.561×10^{-3}
	(0.1, 20)	4.633×10^{-2}	3.824×10^{-2}	4.962×10^{-5}	9.020×10^{-6}
$q = 4$	(0.1, 0.2, 0.3, 0.4)	1.965×10^{-2}	2.953×10^{-2}	2.574×10^{-4}	1.193×10^{-3}
	(0.1, 0.2, 0.3, 20)	1.120×10^{-1}	7.316×10^{-2}	8.364×10^{-5}	9.768×10^{-6}
$q = 8$	(0.1, 0.2, ..., 0.8)	1.241×10^{-2}	9.036×10^{-3}	4.988×10^{-4}	7.841×10^{-4}
	(0.1, ..., 0.7, 20)	1.715×10^{-1}	1.253×10^{-1}	2.134×10^{-4}	9.619×10^{-5}
$q = 10$	(0.1, 0.2, ..., 1.0)	1.067×10^{-2}	6.340×10^{-3}	6.672×10^{-4}	6.580×10^{-4}
	(0.1, ..., 0.9, 20)	1.772×10^{-1}	1.449×10^{-1}	1.745×10^{-4}	1.931×10^{-4}

which has been used in Coelho and Marques [6], where $\phi_W(\cdot)$ and $\phi_{A_i}(\cdot)$ are the characteristic functions of W and A_i , respectively. The measure M_i gives an upper bound of errors of the approximation

$$\sup_{x \in \mathbb{R}^1} |P(W \leq x) - P(A_i \leq x)| \leq M_i.$$

Note that M_i ($i = 1, 2$) measures the global distance between the two distribution functions. We also use the measure based on the difference between them for a particular point $x_{i,\alpha} \in \mathbb{R}^1$, given by

$$M_{i,\alpha} = |P(W \leq x_{i,\alpha}) - P(A_i \leq x_{i,\alpha})|,$$

where $x_{i,\alpha}$ denotes the upper 100α percentile point of A_i , and $P(W \leq x_{i,\alpha})$ is calculated by the numerical integration method described in Imhof [15]. Table 1 shows the values of M_i and $M_{i,\alpha}$ for $\alpha = 0.95$ and the coefficients c_1, \dots, c_q with or without a spiked value. We can see that the global approximation for W with A_1 or A_2 does not work so well, while the pointwise approximation works very well. Although the values of $M_{i,\alpha}$ ($i = 1, 2$) for $\alpha = 0.90, 0.99$ are not listed here, their values are also very small. Comparing the approximations with A_1 and A_2 , it seems that the approximation with A_2 is better when a spiked value is included in the coefficients.

3. Asymptotic distributions for test criteria

In this section, we derive the asymptotic distributions for the test criteria T_D^2 , T_{BS}^2 and T_S^2 , proposed by Dempster [9], Bai and Saranadasa [2] and Srivastava and Du [25] respectively, under H_0 in (1.1) and the condition in (1.7), when both the sample size and the dimension tend to infinity.

3.1. Asymptotic distributions for T_D^2 and T_{BS}^2

We assume that the sample size n tends to infinity when the dimension p tends to infinity, i.e., $n \rightarrow \infty$ as $p \rightarrow \infty$. Let

$$U = \frac{n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n - \text{tr } \Sigma_p}{\sqrt{2\text{tr } \Sigma_p^2}}, \quad V = \frac{(n-1)\text{tr } \mathbf{S}_n - (n-1)\text{tr } \Sigma_p}{\sqrt{2(n-1)\text{tr } \Sigma_p^2}}, \quad (3.1)$$

then the asymptotic null distribution for U under the condition in (1.7) is given by $\Phi(\Sigma_p)$ in (2.3) since $n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n$ has the same distribution as $\mathbf{z}_p^T \mathbf{z}_p$ where \mathbf{z}_p is distributed as $N_p(\mathbf{0}, \Sigma_p)$. Also, the convergence $V \xrightarrow{d} N(0, 1)$ is similarly obtained from the proof of Theorem 2.1. Then we have

$$T_D^2 = 1 + \frac{\sqrt{2\text{tr } \Sigma_p^2}}{\text{tr } \Sigma_p} \{U + O_p(n^{-1/2})\}, \quad T_{BS}^2 = \sqrt{2\text{tr } \Sigma_p^2} \{U + O_p(n^{-1/2})\}. \quad (3.2)$$

Thus, we can get the following theorem.

Theorem 3.1. Under the condition in (1.7) and H_0 , the asymptotic distribution for each of

$$\tilde{T}_D^2 = \frac{\text{tr } \Sigma_p}{\sqrt{2\text{tr } \Sigma_p^2}} \left(\frac{n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n}{\text{tr } \mathbf{S}_n} - 1 \right) \quad \text{and} \quad \tilde{T}_{BS}^2 = \frac{n\bar{\mathbf{X}}_n^T \bar{\mathbf{X}}_n - \text{tr } \mathbf{S}_n}{\sqrt{2\text{tr } \Sigma_p^2}}$$

is given by $\Phi(\Sigma_p)$ defined in (2.3).

Generally, Σ_p is unknown. We need to estimate $\text{tr } \Sigma_p$ and $\text{tr } \Sigma_p^2$. The following theorem gives us an unbiased and ratio-consistent estimator (in the sense that the ratio between true value and its estimator converges to 1 in probability) for each of them, the proof of which is given in Appendix. The estimators of them have been developed by Srivastava [23] and the consistency has been proved under the condition in (1.4). It should be noted that we do not require any conditions for the population covariance matrix in the following theorem.

Theorem 3.2. For all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \sup_{p \geq 0} P \left(\left| \frac{\hat{a}_1}{\text{tr } \Sigma_p} - 1 \right| > \varepsilon \right) = 0, \quad \lim_{n \rightarrow \infty} \sup_{p \geq 0} P \left(\left| \frac{\hat{a}_2}{\text{tr } \Sigma_p^2} - 1 \right| > \varepsilon \right) = 0,$$

where

$$\hat{a}_1 = \text{tr } \mathbf{S}_n, \quad \hat{a}_2 = \frac{(n-1)^2}{(n+1)(n-2)} \left[\text{tr } \mathbf{S}_n^2 - \frac{1}{n-1} (\text{tr } \mathbf{S}_n)^2 \right]. \quad (3.3)$$

As seen in Section 1, T_D^2 is approximated as $F_{r, (n-1)r}$ where r is given in (1.2). Theorem 3.2 shows that the estimator of r is given by

$$\hat{r} = \frac{\hat{a}_1^2}{\hat{a}_2} = \frac{(\text{tr } \mathbf{S}_n)^2}{\frac{(n-1)^2}{(n+1)(n-2)} \left[\text{tr } \mathbf{S}_n^2 - \frac{1}{n-1} (\text{tr } \mathbf{S}_n)^2 \right]},$$

which is clearly ratio consistent.

Remark 3.1. It is noted that $\tilde{T}_D^2 = \tilde{T}_{BS}^2$ when \hat{a}_1 and \hat{a}_2 replace $\text{tr } \Sigma_p$ and $\text{tr } \Sigma_p^2$, respectively. We also note that $\hat{a}_i / \text{tr } \Sigma_p^i = 1 + O_p(n^{-1/2})$, $i = 1, 2$ from the proof of Theorem 3.2.

3.2. Asymptotic distribution for T_S^2

In this section, we assume the condition in (1.7) with \mathcal{R}_p instead of Σ_p , and further assume that not only $n \rightarrow \infty$ as $p \rightarrow \infty$ but also

$$\frac{p^{1-\delta_2}}{n} \rightarrow 0, \quad p \rightarrow \infty, \quad (3.4)$$

where δ_2 is determined by the condition. Note that (3.4) is trivial when $\delta_2 \geq 1$. Let $\Sigma_p(i, j) = \sigma_{ij}$, $\mathbf{S}_n(i, j) = s_{ij}$ and $\mathbf{D}_{S/\sigma} = \text{diag}(s_{11}/\sigma_{11}, \dots, s_{pp}/\sigma_{pp})$, then we have $\mathbf{D}_{S/\sigma}^{-1} = \mathbf{I}_p - \mathbf{D}_1 + \mathbf{D}_2$ where

$$\mathbf{D}_1 = \text{diag} \left(\frac{s_{11}}{\sigma_{11}} - 1, \dots, \frac{s_{pp}}{\sigma_{pp}} - 1 \right),$$

$$\mathbf{D}_2 = \text{diag} \left(\frac{\sigma_{11}}{s_{11}} + \frac{s_{11}}{\sigma_{11}} - 2, \dots, \frac{\sigma_{pp}}{s_{pp}} + \frac{s_{pp}}{\sigma_{pp}} - 2 \right).$$

Define $\mathbf{z}_p = \sqrt{n} \mathbf{D}_{S/\sigma}^{-1/2} \tilde{\mathbf{X}}_n$ where $\mathbf{D}_{S/\sigma} = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$, then \mathbf{z}_p is distributed as $N_p(\mathbf{0}, \mathcal{R}_p)$ under H_0 . Thus, T_S^2 is expressed as

$$T_S^2 = \mathbf{z}_p^T \mathbf{D}_{S/\sigma}^{-1} \mathbf{z}_p = \mathbf{z}_p^T \mathbf{z}_p - \mathbf{z}_p^T \mathbf{D}_1 \mathbf{z}_p + \mathbf{z}_p^T \mathbf{D}_2 \mathbf{z}_p = A - B + C, \quad \text{say,}$$

under H_0 . We note that the asymptotic distribution for $(A - p)/(2 \text{tr } \mathcal{R}_p^2)^{1/2}$ is given by $\Phi(\mathcal{R}_p)$ defined in (2.3).

We next show that each of $\tilde{B} = B/(2 \text{tr } \mathcal{R}_p^2)^{1/2}$ and $\tilde{C} = C/(2 \text{tr } \mathcal{R}_p^2)^{1/2}$ converges to 0 in probability. Following Srivastava and Du [25], we have

$$E(B) = 0, \quad \text{Var}(B) = \frac{6p}{n} + \frac{2}{n} \sum_{i \neq j} (1 + 2r_{ij}^2) r_{ij}^2 \leq \frac{6}{n} \text{tr } \mathcal{R}_p^2,$$

where $\mathcal{R}_p(i, j) = r_{ij}$. Using Chebyshev's inequality, we can show that $\tilde{B} = O_p(n^{-1/2})$. And also we note that C is non-negative definite and so is \mathbf{D}_2 . Then, it is obtained that $E(|\tilde{C}|) = E(C) = 2p/(n-2)$. Thus,

$$E(|\tilde{C}|) = \frac{2p}{n-2} (2 \text{tr } \mathcal{R}_p^2)^{-1/2} = \frac{\sqrt{2} p^{1-\delta_2}}{n-2} \left(\frac{\text{tr } \mathcal{R}_p^2}{p^{2\delta_2}} \right)^{-1/2} = O(p^{1-\delta_2}/n),$$

which implies that $\tilde{C} = O_p(p^{1-\delta_2}/n)$ from Markov's inequality. Hence,

$$\frac{T_S^2 - p}{\sqrt{2 \text{tr } \mathcal{R}_p^2}} = \frac{A - p}{\sqrt{2 \text{tr } \mathcal{R}_p^2}} + O_p \left\{ \max \left(\frac{1}{\sqrt{n}}, \frac{p^{1-\delta_2}}{n} \right) \right\}, \quad (3.5)$$

which leads the following theorem since $E(T_S^2) - p = (n-1)p/(n-3) - p$ converges to 0.

Table 2

Empirical significance level for $\rho = 0.3, 0.5, 0.8$. (N) denotes the approximation with $N(0, 1)$ for the distributions of \tilde{T}_D^2 and \tilde{T}_S^2 , (C) denotes the approximation with $(\chi_1^2 - 1)/\sqrt{2}$ for them..

ρ	p	n	$p^{1-\delta_2}/n$	$T_D^2(N)$	$T_D^2(C)$	$T_S^2(N)$	$T_S^2(C)$
0.3	40	20	0.050	0.0893	0.0686	0.1041	0.0812
	40	40	0.025	0.0769	0.0566	0.0818	0.0626
	100	40	0.025	0.0789	0.0589	0.0848	0.0651
	100	80	0.013	0.0713	0.0540	0.0737	0.0569
	200	40	0.025	0.0844	0.0647	0.0892	0.0717
	200	80	0.013	0.0698	0.0533	0.0731	0.0555
	200	150	0.007	0.0681	0.0515	0.0694	0.0522
	250	250	0.004	0.0663	0.0507	0.0669	0.0517
0.5	40	20	0.050	0.0852	0.0658	0.0934	0.0718
	40	40	0.025	0.0735	0.0562	0.0759	0.0586
	100	40	0.025	0.0791	0.0599	0.0812	0.0631
	100	80	0.013	0.0750	0.0555	0.0758	0.0567
	200	40	0.025	0.0767	0.0560	0.0795	0.0599
	200	80	0.013	0.0761	0.0559	0.0777	0.0578
	200	150	0.007	0.0723	0.0525	0.0732	0.0534
	250	250	0.004	0.0656	0.0488	0.0660	0.0489
0.8	40	20	0.050	0.0912	0.0690	0.0894	0.0682
	40	40	0.025	0.0807	0.0594	0.0795	0.0590
	100	40	0.025	0.0793	0.0585	0.0779	0.0582
	100	80	0.013	0.0735	0.0534	0.0730	0.0532
	200	40	0.025	0.0851	0.0672	0.0849	0.0661
	200	80	0.013	0.0744	0.0545	0.0739	0.0541
	200	150	0.007	0.0698	0.0516	0.0698	0.0513
	250	250	0.004	0.0699	0.0504	0.0696	0.0504

Theorem 3.3. Under H_0 and the conditions in (1.7) with \mathcal{R}_p instead of Σ_p and in (3.4), the asymptotic distribution for

$$\tilde{T}_S^2 = \frac{n\bar{\mathbf{X}}_n^T \mathbf{D}_S^{-1} \bar{\mathbf{X}}_n - \frac{(n-1)p}{n-3}}{\sqrt{2\text{tr } \mathcal{R}_p^2}}$$

is given by $\Phi(\mathcal{R}_p)$ defined in (2.3).

Remark 3.2. In a practical situation, we need to estimate $\text{tr } \mathcal{R}_p^2$. Srivastava and Du [25] have proposed an estimator for $\text{tr } \mathcal{R}_p^2$ given by $\text{tr } \mathbf{R}_n^2 - p^2/(n-1)$ where \mathbf{R}_n denotes the sample correlation matrix. Following Srivastava and Du [25], it can be found that $[\text{tr } \mathbf{R}_n^2 - p^2/(n-1)]/\text{tr } \mathcal{R}_p^2 = 1 + O_p(n^{-1/2})$.

4. Simulations

In Section 3, we have developed the asymptotic distributions for \tilde{T}_D^2 , \tilde{T}_{BS}^2 and \tilde{T}_S^2 under the condition in (1.7). The distributions obtained are not the standard normal distribution but the standardized chi-square distribution, the convolution of normal and chi-squared distributions, and the convolutions of chi-square distributions under the condition. In this section, we conduct several numerical simulations to compare the existing approximation (normal distribution) and the other approximations for the distributions of these criteria in the case where the population covariance (or correlation) matrix does not meet the conditions in (1.3)–(1.5) but meets the condition in (1.7). In view of Remark 3.1, only \tilde{T}_D^2 and \tilde{T}_S^2 are used. We derive actual error probabilities of the first kind for each approximation by Monte Carlo simulations when the significance level is 5% and the number of replications is 10,000. The value of $p^{1-\delta_2}/n$, which depends on the rate of convergence for \tilde{T}_S^2 as shown in (3.5), is also given.

In Tables 2 and 3, we assume $n \times p$ data matrix is generated by

$$X_{ij} = \mu_j + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, p, \quad (4.1)$$

where μ_j 's are constants, $\{\alpha_i\}$ are i.i.d. as $N(0, \sigma_a^2)$, $\{\varepsilon_{ij}\}$ are i.i.d. as $N(0, \sigma_e^2)$ and $\{\alpha_i\}$ and $\{\varepsilon_{ij}\}$ are mutually independent. Let $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$; then $\mathbf{X}_1, \dots, \mathbf{X}_n$ are i.i.d. as $N_p(\mathbf{0}, \Sigma_p)$ under the null hypothesis, where $\Sigma_p = (\sigma_a^2 + \sigma_e^2)[(1-\rho)I_p + \rho \mathbf{1}_p \mathbf{1}_p^T]$ and $\rho = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$. In Table 2, we choose the value of ρ as 0.3, 0.5, 0.8. As seen in Example 2.2, $\delta_2 = 1$ and each of the asymptotic null distributions for \tilde{T}_D^2 and \tilde{T}_S^2 is given by $(\chi_1^2 - 1)/\sqrt{2}$. In Table 2, (N) denotes the approximation with $N(0, 1)$ for the distributions of these statistics, and (C) denotes the approximation with $(\chi_1^2 - 1)/\sqrt{2}$ for them. In Table 3, we set $\sigma_a^2 = p^{-1/2}$ and $\sigma_e^2 = 1$ in the model (4.1). As seen in Example 2.3, $\delta_2 = 1/2$ and each of the asymptotic null distributions

Table 3

Empirical significance level when $\sigma_a^2 = p^{-1/2}$ and $\sigma_e^2 = 1$. (N) denotes the approximation with $N(0, 1)$ for the distributions of \tilde{T}_D^2 and \tilde{T}_S^2 , and (NC) denotes the approximation with $N(0, 1) + (\chi^2 - 1)/\sqrt{2}$ for them.

p	n	$p^{1-\delta_2}/n$	$T_D^2(N)$	$T_D^2(NC)$	$T_S^2(N)$	$T_S^2(NC)$
40	20	0.316	0.0805	0.0724	0.0994	0.0888
40	40	0.158	0.0714	0.0640	0.0809	0.0732
100	40	0.250	0.0692	0.0608	0.0794	0.0689
100	80	0.125	0.0670	0.0591	0.0724	0.0644
200	40	0.354	0.0666	0.0591	0.0726	0.0641
200	80	0.177	0.0613	0.0522	0.0654	0.0566
200	150	0.094	0.0602	0.0522	0.0643	0.0546
250	250	0.063	0.0576	0.0497	0.0590	0.0513

Table 4

Empirical significance level when Σ_p is the block diagonal matrix. (N) denotes the approximation with $N(0, 1)$ for the distributions of \tilde{T}_D^2 and \tilde{T}_S^2 , (I) denotes the numerical integration given in Imhof [15], and (A_1) and (A_2) denote the approximations given by Welch [26] and Pearson [19], respectively.

p	n	$T_D^2(N)$	$T_D^2(I)$	$T_D^2(A_1)$	$T_D^2(A_2)$	$T_S^2(N)$	$T_S^2(I)$	$T_S^2(A_1)$	$T_S^2(A_2)$
40	20	0.0859	0.0639	0.0623	0.0620	0.0961	0.0742	0.0723	0.0720
40	40	0.0806	0.0621	0.0611	0.0608	0.0842	0.0658	0.0650	0.0642
100	40	0.0757	0.0581	0.0567	0.0564	0.0783	0.0605	0.0594	0.0591
100	80	0.0748	0.0559	0.0545	0.0542	0.0762	0.0559	0.0544	0.0541
200	40	0.0746	0.0576	0.0563	0.0562	0.0775	0.0588	0.0577	0.0575
200	80	0.0746	0.0569	0.0558	0.0557	0.0769	0.0580	0.0568	0.0561
200	150	0.0698	0.0512	0.0500	0.0495	0.0713	0.0521	0.0508	0.0507
250	250	0.0679	0.0502	0.0487	0.0483	0.0689	0.0506	0.0491	0.0489
300	300	0.0693	0.0504	0.0490	0.0489	0.0696	0.0501	0.0494	0.0488

for these criteria is the convolution of normal and chi-squared distributions: $N(0, 1) + (\chi^2 - 1)/\sqrt{2}$. In Table 3, (NC) means the approximation with the convolution for the distributions of these statistics.

In Table 4, we assume Σ_p is a block diagonal matrix such that

$$\Sigma_p = \begin{pmatrix} \Sigma_p^{(1)} & 0 \\ 0 & \Sigma_p^{(2)} \end{pmatrix},$$

where $\Sigma_p^{(i)} = (1 - \rho_i)I_{p_i} + \rho_i \mathbf{1}_{p_i} \mathbf{1}_{p_i}^T$ and $p_1 + p_2 = p$, $0 < \rho_i < 1$. We assume $p_i = p/2$ for simplifying the simulation setup and let $\rho_1 = 0.3$ and $\rho_2 = 0.7$. As seen in Example 2.2, $\delta_2 = 1$ and each of the asymptotic distributions of these criteria is given by (2.4) with $c_i = 1$ ($i = 1, 2$) and $\eta = 1/2$. The unknown parameter ρ_i ($i = 1, 2$) in (2.4) is estimated by the maximum likelihood method given by Donner and Koval [11]. The value of $p^{1-\delta_2}/n$ is omitted due to the limited space of the table. To obtain the exact distribution functions of the asymptotic distributions, we use the numerical integration method given in Imhof [15], which is referred to as (I) in Table 4. We also use the two approximations for the asymptotic distributions given by Welch [26] and Pearson [19], which is referred to as (A_1) and (A_2) in Table 4 respectively.

Tables 2–4 show that the new asymptotic distributions work better than the existing asymptotic distribution (normal distribution). It is seen that good approximation requires not only large sample sizes, but also large dimension. Comparing three test criteria, the approximation for T_S^2 is slightly worse than that for the others as shown in Table 3, while all the criteria have almost the same empirical significance level in Tables 2 and 4. The value of $p^{1-\delta_2}/n$ in the rate (3.5) seems not to be small enough. In Table 4, Imhof's method works well for large n and p , while the other two methods work well for small them. It seems that this result is obtained because the 95 percentile points z_1 , z_2 and z_3 of the asymptotic distribution, the Welch's and the Pearson's approximate distributions for each T_D^2 and T_S^2 have the order $z_1 < z_2 < z_3$, and the convergence to the asymptotic distribution is not be attained yet for small n and p .

Remark 4.1. It follows from (3.2), (3.5) and the results in Section 2.2 that for all the situations considered in this section:

$$\begin{aligned} \tilde{T}_D^2 &= \Phi(\Sigma_p) + O_p \left\{ \max(n^{-1/2}, p^{-1/2}) \right\}, \\ \tilde{T}_S^2 &= \Phi(\mathcal{R}_p) + O_p \left\{ \max(n^{-1/2}, n^{-1}p^{1-\delta_2}, p^{-1/2}) \right\}. \end{aligned}$$

Even when one use the estimators of $\text{tr } \Sigma_p$, $\text{tr } \Sigma_p^2$ and $\text{tr } \mathcal{R}_p^2$, it is found that the above rates hold from Remarks 3.1 and 3.2.

The three simulations also show that the results of the tests with the statistics \tilde{T}_D^2 and \tilde{T}_S^2 would not be reliable when the population covariance matrix is misspecified. For this reason, we would need to check the population covariance matrix before testing the mean vector. For instance, see Ledoit and Wolf [18] and Srivastava [23] for testing the population covariance matrix, Jung and Marron [16] and Yata and Aoshima [27] for estimating the parameters in the spiked model (2.5).

Appendix

In this appendix, we give the proofs of [Proposition 2.1](#) and [Theorem 3.2](#).

Proof of Proposition 2.1. When $0 < \alpha < 1/2$, the asymptotic distribution for Z_p is given by $N(0, 1)$. From the well known Berry–Esseen bounds for an independent random sum (see, e.g., Theorem 7.4.1 in Chung [5]), when $0 < \alpha < 1/2$, we have

$$r_p = \frac{c_1^3 p^{3\alpha} + c_2^3 (p-1)}{[c_1^2 p^{2\alpha} + c_2^2 (p-1)]^{3/2}} = O\{\max(p^{3\alpha-3/2}, p^{-1/2})\} = O(p^{-1/2}).$$

For $\alpha = 1/2$, it follows that

$$Z_p = \frac{c_1 p^{1/2} Z_1 - c_1 p^{1/2}}{\sqrt{2c_1^2 + 2c_2^2 (p-1)}} + \frac{\sum_{j=2}^p c_2 U_j - c_2 (p-1)}{\sqrt{2c_1^2 + 2c_2^2 (p-1)}},$$

where $\{U_i\}$ are i.i.d. as χ_1^2 and the second term has asymptotic normality as $p \rightarrow \infty$ whose convergence rate is $p^{-1/2}$. Then, with the random variable N distributed as $N(0, 1)$,

$$Z_p = \frac{c_1 U_1 - c_1}{\sqrt{2(c_1^2 + c_2^2)}} [1 + O(p^{-1})] + \sqrt{\frac{c_2^2}{c_1^2 + c_2^2}} [N + O_p(p^{-1/2})],$$

which leads $r_p = p^{-1/2}$. When $\alpha > 1/2$, note that

$$Z_p = \frac{c_1 p^\alpha U_1 - c_1 p^\alpha}{\sqrt{2c_1^2 p^{2\alpha} + 2c_2^2 (p-1)}} + \frac{\sum_{j=2}^p c_2 U_j - c_2 (p-1)}{\sqrt{2c_1^2 p^{2\alpha} + 2c_2^2 (p-1)}},$$

where the variance of the second term is given by $O(p^{1-2\alpha})$. Thus,

$$Z_p = \frac{U_1 - 1}{\sqrt{2}} [1 + O(p^{1-2\alpha})] + O_p(p^{1/2-\alpha}),$$

which implies $r_p = p^{1/2-\alpha}$. This ends the proof. \square

Before proving [Theorem 3.2](#), we prepare a lemma on some moments of the Wishart distribution. The proof is given by Gupta and Nagar [14] and Srivastava [23] among others.

Lemma A.1. Let \mathbf{W} have a Wishart distribution $W_p(n, \Sigma_p)$. Then,

- (i) $E(\text{tr } \mathbf{W}) = n \text{tr } \Sigma_p$;
- (ii) $\text{Var}(\text{tr } \mathbf{W}) = 2n \text{tr } \Sigma_p^2$;
- (iii) $E(\text{tr } \mathbf{W}^2) = n(n+1) \text{tr } \Sigma_p^2 + n(\text{tr } \Sigma_p)^2$;
- (iv) $E[(\text{tr } \mathbf{W})^2] = 2n \text{tr } \Sigma_p^2 + n^2 (\text{tr } \Sigma_p)^2$;
- (v) $\text{Var}(\text{tr } \mathbf{W}^2) = (8n^3 + 20n^2 + 20n) \text{tr } \Sigma_p^4 + (16n^2 + 16n) \text{tr } \Sigma_p^3 \text{tr } \Sigma_p + (4n^2 + 4n)(\text{tr } \Sigma_p^2)^2 + 8n \text{tr } \Sigma_p^2 (\text{tr } \Sigma_p)^2$;
- (vi) $\text{Var}[(\text{tr } \mathbf{W})^2] = 48 \text{tr } \Sigma_p^4 + 32n^2 \text{tr } \Sigma_p^3 \text{tr } \Sigma_p + 8n^2 (\text{tr } \Sigma_p^2)^2 + 8n^3 \text{tr } \Sigma_p^2 (\text{tr } \Sigma_p)^2$;
- (vii) $\text{Cov}[\text{tr } \mathbf{W}^2, (\text{tr } \mathbf{W})^2] = (24n^2 + 24n) \text{tr } \Sigma_p^4 + (8n^3 + 8n^2 + 16n) \text{tr } \Sigma_p^3 \text{tr } \Sigma_p + 8n (\text{tr } \Sigma_p^2)^2 + 8n^2 \text{tr } \Sigma_p^2 (\text{tr } \Sigma_p)^2$.

Proof of Theorem 3.2. Let $m = n - 1$; then $m\mathbf{S}_n$ is distributed as $W_p(m, \Sigma_p)$. From [Lemma A.1](#), we obtain

$$E(\hat{a}_1) = \text{tr } \Sigma_p, \quad E(\hat{a}_2) = \text{tr } \Sigma_p^2,$$

$$\text{Var}(\hat{a}_1) = \frac{2}{m} \text{tr } \Sigma_p^2,$$

and

$$\text{Var}(\hat{a}_2) = \tau^2 \left[\left(\frac{8}{m} + \frac{20}{m^2} - \frac{28}{m^3} - \frac{48}{m^4} + \frac{48}{m^6} \right) \text{tr } \Sigma_p^4 + \left(\frac{4}{m^2} + \frac{4}{m^3} - \frac{16}{m^4} \right) (\text{tr } \Sigma_p^2)^2 \right]$$

where $\tau = (n-1)^2/[(n+1)(n-2)]$. Using Chebyshev's inequality, we have

$$P\left(\left|\frac{\hat{a}_1}{\text{tr}\Sigma_p} - 1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{\text{Var}(\hat{a}_1)}{(\text{tr}\Sigma_p)^2}, \quad P\left(\left|\frac{\hat{a}_2}{\text{tr}\Sigma_p^2} - 1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{\text{Var}(\hat{a}_2)}{(\text{tr}\Sigma_p^2)^2}.$$

Since

$$\sup_{p \geq 0} \frac{\text{tr}\Sigma_p^2}{(\text{tr}\Sigma_p)^2} \leq 1 \quad \text{and} \quad \sup_{p \geq 0} \frac{\text{tr}\Sigma_p^4}{(\text{tr}\Sigma_p^2)^2} \leq 1,$$

it follows that

$$\lim_{n \rightarrow \infty} \sup_{p \geq 0} \frac{\text{Var}(\hat{a}_1)}{(\text{tr}\Sigma_p)^2} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \sup_{p \geq 0} \frac{\text{Var}(\hat{a}_2)}{(\text{tr}\Sigma_p^2)^2} = 0.$$

The proof of the theorem is completed. \square

References

- [1] G.S. Arunajadai, A point process driven multiple change point model: a robust resistant approach, *Mathematical Biosciences* 220 (2009) 57–71.
- [2] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, *Statistica Sinica* 6 (1996) 311–329.
- [3] M.J. Buckley, G.K. Eagleson, An approximation to the distribution of quadratic forms in normal random variables, *The Australian Journal of Statistics* 30A (1988) 150–159.
- [4] A. Castaño-Martínez, F. López-Blázquez, Distribution of a sum of weighted non central chi-square variables, *TEST* 14 (2005) 397–414.
- [5] K.L. Chung, *A Course in Probability Theory*: Third Edition, Academic press, New York, 2000.
- [6] C.A. Coelho, F.J. Marques, Near-exact distributions for the independence and sphericity likelihood ratio test statistics, *Journal of Multivariate Analysis* 101 (2010) 583–593.
- [7] R.B. Davies, Numerical inversion of a characteristic function, *Biometrika* 60 (1943) 415–417.
- [8] A.W. Davis, A differential equation approach to linear combinations of independent chi-squares, *Journal of the American Statistical Association* 72 (1977) 212–214.
- [9] A.P. Dempster, A high dimensional two sample significance test, *Annals of Mathematical Statistics* 29 (1958) 995–1010.
- [10] A.P. Dempster, A significance test for the separation of two highly multivariate small samples, *Biometrics* 16 (1960) 41–56.
- [11] A. Donner, J.J. Koval, The estimation of intraclass correlation in the analysis of family data, *Biometrics* 36 (1980) 19–25.
- [12] Y. Fujikoshi, T. Himeno, H. Wakaki, Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size, *Journal of the Japan Statistical Society* 34 (2004) 19–26.
- [13] R.M. Gray, Toeplitz and circulant matrices: a review, *Foundations and Trends in Communications and Information Theory* 2 (2006) 155–239.
- [14] A.K. Gupta, D.K. Nagar, *Matrix Variate Distributions*, Chapman and Hall, London, 1999.
- [15] J.P. Imhof, Computing distribution of quadratic forms in normal variables, *Biometrika* 48 (1961) 419–426.
- [16] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, *The Annals of Statistics* 37 (2009) 4104–4130.
- [17] S. Kotz, N.L. Johnson, D.W. Boye, Series representations of distributions of quadratic forms in normal variables I. Central case, *Annals of Mathematical Statistics* 38 (1967) 823–837.
- [18] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, *The Annals of Statistics* 30 (2002) 1081–1102.
- [19] E.S. Pearson, Note on an approximation to the distribution of non-central χ^2 , *Biometrika* 46 (1959) 364.
- [20] F.E. Satterthwaite, Synthesis of variance, *Psychometrika* 6 (1941) 309–316.
- [21] F.E. Satterthwaite, An approximate distribution of estimates of variance components, *Biometrics Bulletin* 2 (1946) 110–114.
- [22] B.K. Shah, C.G. Khatri, Distribution of a definite quadratic form for non-central normal variates, *Annals of Mathematical Statistics* 32 (1961) 883–887.
- [23] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, *Journal of the Japan Statistical Society* 35 (2005) 251–272.
- [24] M.S. Srivastava, Multivariate theory for analyzing high dimensional data, *Journal of the Japan Statistical Society* 37 (2007) 53–86.
- [25] 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.
- [26] B.L. Welch, The significance of the difference between two means when the population variances are unequal, *Biometrika* 29 (1938) 350–361.
- [27] K. Yata, M. Aoshima, Intrinsic dimensionality estimation of high-dimension, low sample size data with d -asymptotics, *Communications in Statistics—Theory and Methods* 39 (2010) 1511–1521.