Scandinavian Journal of Statistics
Theory and Applications

Scandinavian Journal of Statistics, Vol. 42: 988–1009, 2015
doi: 10.1111/sjos.12147


# Likelihood Ratio Tests for High-Dimensional Normal Distributions

TIEFENG JIANG
*School of Statistics, University of Minnesota*

YONGCHENG QI
*Department of Mathematics and Statistics, University of Minnesota*

**ABSTRACT. In their recent work, Jiang and Yang studied six classical Likelihood Ratio Test statistics under high-dimensional setting. Assuming that a random sample of size $n$ is observed from a $p$-dimensional normal population, they derive the central limit theorems (CLTs) when $p$ and $n$ are proportional to each other, which are different from the classical chi-square limits as $n$ goes to infinity, while $p$ remains fixed. In this paper, by developing a new tool, we prove that the mentioned six CLTs hold in a more applicable setting: $p$ goes to infinity, and $p$ can be very close to $n$. This is an almost sufficient and necessary condition for the CLTs. Simulations of histograms, comparisons on sizes and powers with those in the classical chi-square approximations and discussions are presented afterwards.**

*Key words:* central limit theorem, covariance matrix, high-dimensional data, hypothesis test, likelihood ratio test, mean vector, multivariate Gamma function, multivariate normal distribution

## 1. Introduction

Analyzing data with large dimensionality of the population and large sample size is one of the very active areas in mathematical sciences. This is particularly true in Statistics. If $n$ points from a population with dimension $p$ are sampled and are put together, we then see an $n \times p$ matrix naturally. When the population is the multivariate normal distribution, the methodology of studying such data is elaborated in the field of Multivariate Analysis.

In the last decade, with the development and improvement of modern technologies. such as the speed of computers, biology, Wall Street trading and weather forecast, the collected data have a common feature in which both $n$ and $p$ are very large. Thus, renovating the old statistical methods and creating new methods are necessary. There are some recent literatures about these development. For example, in the field of multivariate analysis, Schott (2001, 2005, 2007), Ledoit & Wolf (2002), Bai *et al.* (2009), Chen *et al.* (2010), Jiang *et al.* (2012), and Jiang & Yang (2013) study the classical Likelihood Ratio Tests (LRTs) when $p$ is large. For literatures on large $n$ and large $p$ with other interests, see, for example, two book-length treatments by Serdobolskii (2010) and Fujikoshi *et al.* (2010).

In this paper, we will investigate a problem asked by Jiang & Yang (2013). Our solutions become very applicable in practice. To make the problem clear to understand, let us take one example to illustrate. For a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ is mean vector and $\boldsymbol{\Sigma}$ is $p \times p$ covariance matrix, consider the following spherical test:

$$H_0 : \boldsymbol{\Sigma} = \lambda \mathbf{I}_p \;\; \text{vs} \;\; H_a : \boldsymbol{\Sigma} \neq \lambda \mathbf{I}_p \tag{1.1}$$

with unspecified $\lambda$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent and identically distributed (iid) $\mathbb{R}^p$-valued random vectors with normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $V_n$ be the LRT statistic (given in (2.2)).

The traditional theory of the multivariate analysis says that $-n \log V_n$ goes to a chi-square distribution as $n$ tends to infinity, while $p$ is fixed. Jiang & Yang (2013) prove that it is no longer true as $p \to \infty$. In fact, one of their results show that the central limit theorem (CLT) holds, that is, $(\log V_n - \mu_n)/\sigma_n$ actually converges to the standard normal distribution as $n \to \infty$ and $p/n \to y \in (0, 1]$, where $\mu_n$ and $\sigma_n$ are explicit constants of $n$ and $p$. Similar results on other five classical LRTs are also obtained in their paper. By a comparison between the histograms of $(\log V_n - \mu_n)/\sigma_n$ and the standard normal curve, they observe that the mentioned CLT also holds even when $p$ is large but not necessarily proportional to $n$, that is, the assumption $p/n \to y \in (0, 1]$ does not have to hold. See the comment in problem 3 in section 4 from their paper. Of course, if this is true, the CLT will be very useful in practice because it is hard to judge whether $p/n$ has a limit in $(0, 1]$ for real data.

We prove, in this paper, that the mentioned CLT holds when $p$, which has to be less than $n$ in the LRT, is large but not necessarily at the same scale of $n$. Other five classical LRTs are also shown to have similar behaviours. Therefore, in the corresponding LRTs, the sizes of data are allowed to be more flexible. One does not need to concern if the value of $p$ is large enough to be comparable with the sample size.

One of the key reasons to assume $p/n \to (0, 1]$ in the previous studies relies on the fact that it is a very typical assumption in the field of random matrix theory. In their enlightening work, Bai *et al.* (2009) study a test similar to (1.1) by using the CLT of the eigenvalues of the Wishart matrices. Their work is based on the assumption $p/n \to y \in (0, 1]$. The subsequent work aforementioned naturally use this condition. In this paper, we develop a machinery (proposition 5.1) to deal with the case when $p$ is much smaller than $n$. It is an expansion of the generalized gamma function $\Gamma_p(z)$ (defined by (5.1)) that enables us to obtain the CLT as long as $p \to \infty$ regardless of the relative speed to $n$. Our starting step is the method of moment generating functions. When changing the point from the random matrix theory to the method of the moment generating function, it is very interesting to see that the CLT actually holds for such a big range of $n$ and $p$.

The difference between this work and the work of Jiang & Yang (2013) is as follows: Jiang & Yang study six classical LRTs under the assumption $p/n \to y \in (0, 1]$ or similar conditions for several normal distributions. In this paper, we study the same six tests under the condition $n - c > p \to \infty$ for some $1 \le c \le 4$. So the earlier work is a special case of the current one. In fact, the assumption that '$n - c > p \to \infty$' is almost necessary: when $p$ is finite, the test statistic converges to a chi-square distribution by a classical LRT theorem; when $n - 1 < p$, the LRT does not exists (see the work of Jiang & Yang (2013) for further details). Second, the new results are more applicable. Lastly, the derivation of our new tool of proposition 5.1 is more challenging than lemma 5.4 in the work of Jiang & Yang (2013). Both are the core steps in the proofs appearing in the two papers. Readers are referred to their paper for more descriptions and narrations about the six tests.

The outline of the rest of this paper is given as follows: We present the six LRTs in sections 2.1–2.6. They are the following: (i) testing covariance matrices of normal distributions proportional to identity matrix; (ii) testing independence of components of normal distributions; (iii) testing multiple normal distributions being identical; (iv) testing equality of several covariance matrices; (v) testing specified values for mean vector and covariance matrix; and (vi) testing complete independence of a normal distribution. The CLTs are presented in those sections. In section 3.1, we make pictures to compare the classical chi-square approximations with our CLTs. In section 3.2, we give tables on sizes and powers of the tests. In Section 4, we provide a summary and lead some discussions. Theorem 6 is proved in section 5. The other five theorems are proved in the Supplement by Jiang & Qi (2015).

## 2. Main results

Throughout the paper, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the $p$-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. For any given $\alpha \in (0, 1)$, $\chi^2_{f,\alpha}$ denotes the $\alpha$ level critical value of $\chi^2_f$, the chi-square random variable or the chi-square distribution with $f$ degrees of freedom and $z_\alpha$ denotes the $\alpha$ level critical value of the standard normal distribution $N(0, 1)$. The notation $|\mathbf{A}|$ or $\det(\mathbf{A})$ stands for the determinant of the square matrix $\mathbf{A}$.

In this section, we present the CLTs of six classical LRT statistics. The six theorems are stated in six subsections.

### *2.1. Testing covariance matrices of normal distributions proportional to identity matrix*

Consider a normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The spherical test is given by

$$H_0 : \boldsymbol{\Sigma} = \lambda \mathbf{I}_p \ \text{ vs } \ H_a : \boldsymbol{\Sigma} \neq \lambda \mathbf{I}_p \tag{2.1}$$

with unspecified $\lambda$. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are iid $\mathbb{R}^p$-valued random vectors with normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. As usual, set

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \ \text{ and } \ \mathbf{S} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'.$$

Mauchly (1940) shows that the LRT statistic of (2.1) is as follows:

$$V_n = |\mathbf{S}| \cdot \left(\frac{\operatorname{tr}(\mathbf{S})}{p}\right)^{-p}. \tag{2.2}$$

In this paper, we have the following result about $V_n$.

**Theorem 1.** *Let $p = p_n$ such that $n > p + 1$ for all $n \geq 3$ and $V_n$ be as in (2.2). Assume $\lim_{n\to\infty} p_n = \infty$, then, under $H_0$ in (2.1), $(\log V_n - \mu_n)/\sigma_n$ converges in distribution to $N(0, 1)$ as $n \to \infty$, where*

$$\mu_n = -p - \left(n - p - \frac{3}{2}\right)\log\left(1 - \frac{p}{n-1}\right) \ \text{ and }$$
$$\sigma_n^2 = -2\left[\frac{p}{n-1} + \log\left(1 - \frac{p}{n-1}\right)\right].$$

Jiang & Yang (2013) prove the mentioned theorem for the special case $p/n \to y \in (0, 1]$. Theorem 1 says that $p/n$ does not need to have a limit. Furthermore, if $\lim p/n = y$ exists, the theorem holds for $y = 0$. Theorem 1 holds as long as $p \to \infty$ and $p < n - 1$. It is explained in the work of Jiang & Yang (2013) that the LRT statistic does not exist if $p > n - 1$.

If $p$ is fixed, the classical chi-square approximation says that

$$-(n-1)\rho \log V_n \ \text{ converges to } \ \chi^2_f \tag{2.3}$$

in distribution as $n \to \infty$, where

$$\rho = 1 - \frac{2p^2 + p + 2}{6(n-1)p} \ \text{ and } \ f = \frac{1}{2}(p-1)(p+2).$$

See, for example, the work of Muirhead (1982) or the summary in the work of Jiang & Yang (2013).

Let $\alpha \in (0, 1)$ be any given number. Recall that a LRT of size-$\alpha$ rejects the null hypotheses if the likelihood ratio (or any of its monotone increasing functions) is smaller than a constant $c_\alpha$ that is chosen in such a way that the size or type I error is (approximately) equal to the given $\alpha$. Therefore, the rejection region of LRT of (2.1) is $V_n \le c_\alpha$. Based on the chi-square approximation in (2.3), an approximate size-$\alpha$ rejection region is $-(n-1)\rho \log V_n \ge \chi^2_{f,\alpha}$. Based on the normal approximation in theorem 1, the rejection region is $(\log V_n - \mu_n)/\sigma_n \le -z_\alpha$.

### 2.2. *Testing independence of components of normal distributions*

For $k \ge 2$, let $p_1, \ldots, p_k$ be $k$ positive integers. Denote $p = p_1 + \cdots + p_k$, and let

$$\mathbf{\Sigma} = (\mathbf{\Sigma}_{ij})_{1 \le i,j \le k}$$

be a positive definite matrix, where $\mathbf{\Sigma}_{ij}$ is a $p_i \times p_j$ sub-matrix for all $1 \le i, j \le k$. Assume $\boldsymbol{\xi}_i$ is a $p_i$-variate normal random vector for each $1 \le i \le k$, and $\left(\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_k'\right)'$ has the distribution $N_p(\mu, \mathbf{\Sigma})$. We are interested in testing the independence of $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$, or equivalently testing

$$H_0 : \mathbf{\Sigma}_{ij} = \mathbf{0} \text{ for all } 1 \le i < j \le k \text{ versus } H_a : H_0 \text{ is not true.} \tag{2.4}$$

Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are iid from distribution $N_p(\mu, \mathbf{\Sigma})$. Set $n = N - 1$. Define

$$\mathbf{A} = \sum_{i=1}^{n+1} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \quad \text{with } \bar{\mathbf{x}} = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbf{x}_i,$$

and partition it as follows:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1k} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2k} \\ \vdots & \cdots & \cdots & \vdots \\ \mathbf{A}_{k1} & \mathbf{A}_{k2} & \cdots & \mathbf{A}_{kk} \end{pmatrix},$$

where $\mathbf{A}_{ij}$ is a $p_i \times p_j$ matrix. The likelihood ratio statistic for testing (2.4) is given by

$$\Lambda_n = \frac{|\mathbf{A}|^{(n+1)/2}}{\prod_{i=1}^k |\mathbf{A}_{ii}|^{(n+1)/2}} := (W_n)^{(n+1)/2}, \tag{2.5}$$

see the work of Wilks (1935) or theorem 11.2.1 from the work of Muirhead (1982).

Because $\log \Lambda_n = \frac{n+1}{2} \log W_n$ from (2.5), it suffices to deal with the limiting distribution of $\log W_n$. Assume $p_i := p_{i,n}$ for all $1 \le i \le k$. We have the following result for $\log W_n$:

**Theorem 2.** *Suppose that $n > p$ for all large $n$, and there exists $\delta \in (0, 1)$ satisfying $\delta \le p_i/p_j \le \delta^{-1}$ for all $1 \le i, j \le k$ and all large $n$. Recall $W_n$ as defined in (2.5). If $\min_{1 \le i \le k} p_i \to \infty$ as $n \to \infty$, then, under $H_0$ in (2.4), $(\log W_n - \mu_n)/\sigma_n$ converges in distribution to $N(0, 1)$ where*

$$\mu_n = -r_n^2\left(p - n + \frac{1}{2}\right) + \sum_{i=1}^k r_{n,i}^2 \left(p_i - n + \frac{1}{2}\right) \text{ and } \sigma_n^2 = 2r_n^2 - 2\sum_{i=1}^k r_{n,i}^2;$$

$r_x = \left(-\log\left(1 - \frac{p}{x}\right)\right)^{1/2}$ *for* $x > p$; *and* $r_{x,i} = \left(-\log\left(1 - \frac{p_i}{x}\right)\right)^{1/2}$ *for* $x > p_i$ *and* $1 \le i \le k$.

The assumption '$\delta \le p_i/p_j \le \delta^{-1}$ for all $1 \le i, j \le n$ and all $n$' requires that the sizes of the components $p_i$'s are comparable. This rules out the unusual situation that some of the $p_i$'s are much larger than the others. As was pointed out by Jiang & Yang (2013), the LRT fails if $p > N = n + 1$ because the matrix $\mathbf{A}$ is not of full rank. Jiang & Yang (2013) prove the mentioned theorem under condition that $\lim_{n\to\infty} p_i/n = y_i \in (0, 1]$ for $1 \le i \le k$. When $p_1, p_2, \ldots, p_k$ are fixed as $n$ goes to infinity, the classical LRT statistic of (2.4) has a chi-square limit:

$$-2\rho \log \Lambda_n \text{ converges to } \chi^2_f \tag{2.6}$$

in distribution, where

$$f = \frac{1}{2}\left(p^2 - \sum_{i=1}^{k} p_i^2\right) \text{ and } \rho = 1 - \frac{2\left(p^3 - \sum_{i=1}^k p_i^3\right) + 9\left(p^2 - \sum_{i=1}^k p_i^2\right)}{6(n+1)\left(p^2 - \sum_{i=1}^k p_i^2\right)};$$

see, for example, theorem 11.2.5 in the work of Muirhead (1982).

Let $\alpha \in (0, 1)$ be any given number. Based on the chi-square approximation in (2.6), the LRT rejects the null hypothesis in (2.4) if $-2\rho \log \Lambda_n \ge \chi^2_{f,\alpha}$. Based on the normal approximation in theorem 2, the rejection region is $(\log W_n - \mu_n)/\sigma_n \le -z_\alpha$.

### 2.3. *Testing that multiple normal distributions are identical*

Consider normal distributions $N_p(\mu_i, \mathbf{\Sigma}_i)$, $i = 1, 2, \ldots, k$, where $k \ge 2$. We are interested in testing whether the $k$ distributions are identical, that is,

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k,\ \mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k \text{ versus } H_a : H_0 \text{ is not true.} \tag{2.7}$$

Assume $\{\mathbf{y}_{ij};\ 1 \le i \le k,\ 1 \le j \le n_i\}$ are independent $p$-dimensional random vectors, and for each $i = 1, 2, \ldots, k$, $\{\mathbf{y}_{ij};\ 1 \le j \le n_i\}$ are iid from $N(\mu_i, \mathbf{\Sigma}_i)$. Define the following:

$$\mathbf{A} = \sum_{i=1}^{k} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})', \qquad \mathbf{B}_i = \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)' \text{ and}$$
$$\mathbf{B} = \sum_{i=1}^{k} \mathbf{B}_i = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)',$$

where

$$\bar{\mathbf{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{y}_{ij},\ \ \bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{k} n_i\bar{\mathbf{y}}_i,\ \ n = \sum_{i=1}^{k} n_i.$$

The LRT statistic for (2.7) is first derived by Wilks (1932) as follows:

$$\Lambda_n = \frac{\prod_{i=1}^k |\mathbf{B}_i|^{n_i/2}}{|\mathbf{A} + \mathbf{B}|^{n/2}} \cdot \frac{n^{pn/2}}{\prod_{i=1}^k n_i^{pn_i/2}}. \tag{2.8}$$

See also theorem 10.8.1 from the work of Muirhead (1982). It is noted in the work of Jiang & Yang (2013) that, when $p \ge n_i$ for any $i = 1, \ldots, k$, the determinant of the matrix $\mathbf{B}_i$ is zero because $\mathbf{B}_i$ is not of full rank, and consequently, the likelihood ratio statistic $\Lambda_n$ is zero. Thus, the condition $p < \min\{n_i;\ 1 \le i \le k\}$ is required to ensure that the LRT statistic for the test (2.7) is non-degenerate.

We have the following result for the limiting distribution of $\Lambda_n$ defined in (2.8).

**Theorem 3.** *Let $n_i = n_i(p) > p + 1$ for all $p$, and there exists $\delta \in (0, 1)$ such that $\delta < n_i/n_j < \delta^{-1}$ for all $1 \le i, j \le k$. Let $\Lambda_n$ be as in (2.8). Then, under $H_0$ in (2.7),*

$$\frac{\log \Lambda_n - \mu_n}{n\sigma_n} \text{ converges to } N(0,1)$$

*in distribution as $p \to \infty$, where*

$$\mu_n = \frac{1}{4}\Big[-2kp - \sum_{i=1}^{k} \frac{p}{n_i} + nr_n^2(2p - 2n + 3) - \sum_{i=1}^{k} n_i r_{n_i'}^2 (2p - 2n_i + 3)\Big],$$

$$\sigma_n^2 = \frac{1}{2}\left(\sum_{i=1}^{k} \frac{n_i^2}{n^2} r_{n_i'}^2 - r_n^2\right) > 0,$$

*$n_i' = n_i - 1$ and $r_x = \left(-\log\left(1 - \frac{p}{x}\right)\right)^{1/2}$ for $x > p$.*

If the dimension $p$ is fixed and the null hypothesis in (2.7) is true, it follows, from theorem 10.8.4 in the work of Muirhead (1982), that

$$-2\rho \log \Lambda_n \text{ converges to } \chi_f^2 \tag{2.9}$$

in distribution as $\min_{1 \le i \le k} n_i \to \infty$, where

$$f = \frac{1}{2}p(k-1)(p+3) \text{ and } \rho = 1 - \frac{2p^2 + 9p + 11}{6(k-1)(p+3)n}\left(\sum_{i=1}^{k} \frac{n}{n_i} - 1\right).$$

When $p$ grows with the same rate of $n_i$, namely, $\lim_{p\to\infty} p/n_i = y_i \in (0, 1]$ for $1 \le i \le k$, the mentioned theorem is proved by Jiang & Yang (2013). We should mention that $\mu_n$ in theorem 3 in the work of Jiang & Yang (2013) is slightly defined differently: the counterpart of $\sum_{i=1}^{k} y_i$ in their result is $\sum_{i=1}^{k} \frac{p}{n_i}$ in the mentioned theorem. Note that, in our theorem 3, we do not assume the limits of $\frac{p}{n_i}$ exist. This substitution does not change the limiting distribution because both $\sum_{i=1}^{k} \frac{p}{n_i}$ and $\sum_{i=1}^{k} y_i$ are bounded by $k$; therefore, they are negligible compared with $n\sigma_n$, which converges to infinity as shown in the proof of theorem 3.

Let $\alpha \in (0, 1)$ be any given number. Based on the chi-square approximation in (2.9), the LRT rejects the null hypothesis in (2.7) if $-2\rho \log \Lambda_n \ge \chi^2_{f,\alpha}$. Based on our normal approximation, the rejection region is $(\log \Lambda_n - \mu_n)/(n\sigma_n) \le -z_\alpha$.

### 2.4. *Testing equality of several covariance matrices*

Let $k \ge 2$ be an integer. Consider $p$-dimensional normal distributions $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $1 \le i \le k$, where $\mu_i$ and $\boldsymbol{\Sigma}_i$ are unknown. We are interested in testing

$$H_0 : \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k \quad \text{versus} \quad H_a : H_0 \text{ is not true.} \tag{2.10}$$

For $1 \le i \le k$, let $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$ be iid $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$-distributed random vectors. Define the following:

$$\bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_{ij}, \quad \mathbf{A}_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)', \quad 1 \le i \le k,$$

and

$$\mathbf{A} = \mathbf{A}_1 + \cdots + \mathbf{A}_k \quad \text{and} \quad n = n_1 + \cdots + n_k.$$

The LRT statistic for (2.10), derived in the work of Wilks (1932), is given by

$$\Lambda_n = \frac{\prod_{i=1}^{k} (|\mathbf{A}_i|)^{n_i/2}}{(|\mathbf{A}|)^{n/2}} \cdot \frac{n^{np/2}}{\prod_{i=1}^{k} n_i^{n_i p/2}}.$$

The test rejects the null hypothesis $H_0$ when $\Lambda_n \le c_\alpha$, where $c_\alpha$ is selected such that the test has the significance level of $\alpha$. The test statistic $\Lambda_n$ is non-degenerate only if all determinants of $\mathbf{A}_i$ are nonzero; hence, the condition that $p < n_i$ for all $i = 1, \ldots, k$ is required. We are interested in the following modified LRT statistic $\Lambda_n^*$, suggested by Bartlett (1937):

$$\Lambda_n^* = \frac{\prod_{i=1}^{k} (|\mathbf{A}_i|)^{(n_i-1)/2}}{(|\mathbf{A}|)^{(n-k)/2}} \cdot \frac{(n-k)^{(n-k)p/2}}{\prod_{i=1}^{k} (n_i-1)^{(n_i-1)p/2}}. \tag{2.11}$$

This modified LRT has been proved to be unbiased, for example, in the works of Sugiura & Nagao (1968) and Perlman (1980). In this paper, we will prove the following CLT for $\log \Lambda_n^*$.

**Theorem 4.** *Assume $n_i = n_i(p)$ for all $1 \le i \le k$ such that $\min_{1\le i\le k} n_i > p+1$, and there exists $\delta \in (0,1)$ such that $\delta < n_i/n_j < \delta^{-1}$ for all $i, j$. Let $\Lambda_n^*$ be as in (2.11). Then, under $H_0$ in (2.10), $\left(\log \Lambda_n^* - \mu_n\right)/((n-k)\sigma_n)$ converges weakly to $N(0,1)$ as $p \to \infty$, where*

$$\begin{aligned}\mu_n = \frac{1}{4}\Bigg[&(n-k)(2n-2p-2k-1)\log\left(1-\frac{p}{n-k}\right)\\ &-\sum_{i=1}^{k}(n_i-1)(2n_i-2p-3)\log\left(1-\frac{p}{n_i-1}\right)\Bigg] \text{ and}\\ \sigma_n^2 = \frac{1}{2}\Bigg[&\log\left(1-\frac{p}{n-k}\right)-\sum_{i=1}^{k}\left(\frac{n_i-1}{n-k}\right)^2\log\left(1-\frac{p}{n_i-1}\right)\Bigg] > 0.\end{aligned}$$

The CLT for $\log \Lambda_n^*$ has also been studied by Bai *et al.* (2009), Jiang *et al.* (2012) and Jiang & Yang (2013). Jiang & Yang (2013) prove theorem 4 under more restrictive condition that $p/n_i \to y_i \in (0,1]$ for $i = 1, \ldots, k$. When $p$ is fixed, the classical chi-square approximation is obtained by Box (1949). Under the null hypothesis of (2.10), Box (1949) shows that

$$-2\rho \log \Lambda_n^* \text{ converges to } \chi_f^2 \tag{2.12}$$

in distribution as $\min_{1\le i\le k} n_i \to \infty$, where

$$f = \frac{1}{2}p(p+1)(k-1) \text{ and } \rho = 1 - \frac{2p^2+3p-1}{6(p+1)(k-1)(n-k)}\left(\sum_{i=1}^{k}\frac{n-k}{n_i-1}-1\right).$$

Let $\alpha \in (0,1)$ be any given number. Based on the chi-square approximation, the LRT rejects the null hypothesis in (2.10) if $-2\rho\log\Lambda_n^* \ge \chi^2_{f,\alpha}$. Based on the normal approximation in theorem 4, the rejection region is $\left(\log\Lambda_n^* - \mu_n\right)/((n-k)\sigma_n) \le -z_\alpha$.

### *2.5. Testing specified values for mean vector and covariance matrix*

Consider a normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean vector and $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix. Based on $n$ iid random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from the normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we test

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ and } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \text{ versus } H_a : H_0 \text{ is not true},$$

where $\boldsymbol{\mu}_0$ is a given vector in $\mathbb{R}^p$, and $\boldsymbol{\Sigma}_0$ is a given $p \times p$ non-singular matrix. Through the data transformation $\tilde{\mathbf{x}}_i = \boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0)$, the mentioned hypothesis test is equivalent to the test

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \text{ and } \boldsymbol{\Sigma} = \mathbf{I}_p \quad \text{versus} \quad H_a : H_0 \text{ is not true.} \tag{2.13}$$

Set

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \quad \mathbf{A} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'.$$

The LRT statistic of (2.13) is given by

$$\Lambda_n = \left(\frac{e}{n}\right)^{np/2} |\mathbf{A}|^{n/2} e^{-\mathrm{tr}(\mathbf{A})/2} e^{-n\bar{\mathbf{x}}'\bar{\mathbf{x}}/2}; \tag{2.14}$$

for example, theorem 8.5.1 in the work of Muirhead (1982). The condition that $p < n$ is required to ensure that $\Lambda_n$ is non-degenerate. We have the following CLT for $\log \Lambda_n$:

**Theorem 5.** *Assume that $p := p_n$ such that $n > 1 + p$ for all $n \geq 3$ and $p \to \infty$ as $n$ goes to infinity. Let $\Lambda_n$ be defined as in (2.14). Then, under the null hypothesis of (2.13),* $(\log \Lambda_n - \mu_n)/(n\sigma_n)$ *converges in distribution to $N(0,1)$ as $n \to \infty$, where*

$$\mu_n = -\frac{1}{4}\left[n(2n-2p-3)\log\Big(1 - \frac{p}{n-1}\Big) + 2(n+1)p\right] \quad \textit{and}$$
$$\sigma_n^2 = -\frac{1}{2}\left(\frac{p}{n-1} + \log\Big(1 - \frac{p}{n-1}\Big)\right) > 0.$$

Jiang & Yang (2013) show theorem 5 under a stronger condition, $p/n \to y \in (0,1]$. When $p$ is fixed, it follows from theorem 8.5.5 in the work of Muirhead (1982) that, as $n \to \infty$,

$$-2\rho \log \Lambda_n \quad \text{converges to} \quad \chi^2_f \tag{2.15}$$

under the null hypothesis of (2.13), where

$$\rho = 1 - \frac{2p^2 + 9p + 11}{6n(p+3)} \quad \text{and} \quad f = \frac{1}{2}p(p+3).$$

Let $\alpha \in (0,1)$ be any given number. Based on the chi-square approximation in (2.15), the LRT rejects the null hypothesis in (2.13) if $-2\rho \log \Lambda_n \geq \chi^2_{f,\alpha}$. Based on the normal approximation in theorem 5, the rejection region is $(\log \Lambda_n - \mu_n)/\sigma_n \leq -z_\alpha$.

### *2.6. Testing complete independence*

Assume that a $p$-dimensional random vector $\mathbf{x} = (x_1, \ldots, x_p)'$ has a distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We are interested in testing that the $p$ components $x_1, x_2, \ldots, x_p$ are independent or equivalently testing that the covariance matrix $\boldsymbol{\Sigma}$ is diagonal. Let $\mathbf{R} = (r_{ij})_{p \times p}$ be the correlation matrix generated from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the test can be written as follows:

$$H_0 : \mathbf{R} = \mathbf{I}_p \quad \text{versus} \quad H_a : \mathbf{R} \neq \mathbf{I}_p. \tag{2.16}$$

To obtain the asymptotic distribution for the test statistic of the LRT for (2.16), we will consider a larger class of distributions, namely, *spherical distributions*. Recall that a random vector $\mathbf{y} \in \mathbb{R}^n$ has a *spherical distribution* if $\mathbf{O}\mathbf{y}$ and $\mathbf{y}$ have the same probability distribution for all $n \times n$ orthogonal matrix $\mathbf{O}$. Obviously, any $n$-dimensional normal random vector with distribution $N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ has a *spherical distribution* for any $\sigma > 0$.

Let $\mathbf{X} = (x_{ij})_{n\times p} = (\mathbf{x}_1, \dots, \mathbf{x}_n)' = (\mathbf{y}_1, \dots, \mathbf{y}_p)$ be an $n \times p$ matrix such that $\mathbf{y}_1, \dots, \mathbf{y}_p$ are independent random vectors with $n$-variate spherical distributions, and $P(\mathbf{y}_i = \mathbf{0}) = 0$ for all $1 \le i \le p$ (these distributions may be different). For $1 \le i, j \le p$, let $\hat{r}_{ij}$ denote the Pearson correlation coefficient between $(x_{1i}, \dots, x_{ni})$ and $(x_{1j}, \dots, x_{nj})$, given by

$$\hat{r}_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \cdot \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}, \tag{2.17}$$

where $\bar{x}_i = \frac{1}{n}\sum_{k=1}^n x_{ki}$ and $\bar{x}_j = \frac{1}{n}\sum_{k=1}^n x_{kj}$. Then

$$\hat{\mathbf{R}}_n := (\hat{r}_{ij})_{p\times p} \tag{2.18}$$

is the sample correlation matrix based on the $p$-dimensional random vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$. A natural requirement for non-singularity of $\hat{\mathbf{R}}_n$ is $n > p$. From theorem 5.1.3 in the work of Muirhead (1982), the density of $|\hat{\mathbf{R}}_n|$ is given by

$$\text{Constant} \cdot |\mathbf{R}_n|^{(n-p-2)/2} d\mathbf{R}_n. \tag{2.19}$$

We first present the limiting distribution concerning the determinant of $\hat{\mathbf{R}}_n$.

**Theorem 6.** *Let $p = p_n$ satisfy that $n > p + 4$ and $p \to \infty$. Let $\mathbf{X} = (\mathbf{y}_1, \dots, \mathbf{y}_p)$ be an $n \times p$ matrix such that $\mathbf{y}_1, \dots, \mathbf{y}_p$ are independent random vectors with $n$-variate spherical distribution, and $P(\mathbf{y}_i = \mathbf{0}) = 0$ for all $1 \le i \le p$ (these distributions may be different). Then* $(\log|\hat{\mathbf{R}}_n| - \mu_n)/\sigma_n$ *converges in distribution to $N(0,1)$ as $n \to \infty$, where*

$$\mu_n = \left(p - n + \frac{3}{2}\right)\log\left(1 - \frac{p}{n-1}\right) - \frac{n-2}{n-1}p,$$
$$\sigma_n^2 = -2\left[\frac{p}{n-1} + \log\left(1 - \frac{p}{n-1}\right)\right].$$

The theorem is proved by Jiang & Yang (2013) under the condition that $n > p + 4$ and $p/n \to y \in (0, 1]$.

Now, we return to the LRT of (2.16). Assume random vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ are iid from a $p$-variate normal distribution $N_p(\mu, \mathbf{\Sigma})$ with a correlation matrix $\mathbf{R}$. Write $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ for $1 \le i \le n$, and define the sample correlation matrix $\hat{\mathbf{R}}_n = (\hat{r}_{ij})_{p\times p}$ as in (2.17). From the work of Morrison (2005), page 40, the rejection region of the LRT for (2.16) is

$$|\hat{\mathbf{R}}_n|^{n/2} \le c_\alpha,$$

where $c_\alpha$ is determined so that the test has significance level of $\alpha$. To derive the asymptotic distribution of $\log|\hat{\mathbf{R}}_n|$, write $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)' = (x_{ij})_{n\times p} = (\mathbf{y}_1, \dots, \mathbf{y}_p)$. Then, under the null hypothesis of (2.16), $\mathbf{y}_1, \dots, \mathbf{y}_p$ are independent random vectors from $n$-variate normal distributions (these normal distributions may differ by their covariance matrices), and $P(\mathbf{y}_i = 0) = 0$ for all $1 \le i \le p$. Therefore, we have the following corollary to theorem 6:

**Corollary 1.** *Assume that $p := p_n$ satisfy that $n > p + 4$ and $p \to \infty$. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be iid from $N_p(\mu, \mathbf{\Sigma})$ with the Pearson sample correlation matrix $\hat{\mathbf{R}}_n$ as defined in (2.18). Then, under $H_0$ in (2.16),* $(\log|\hat{\mathbf{R}}_n| - \mu_n)/\sigma_n$ *converges in distribution to $N(0,1)$ as $n \to \infty$, where $\mu_n$ and $\sigma_n$ are given as in theorem 6.*

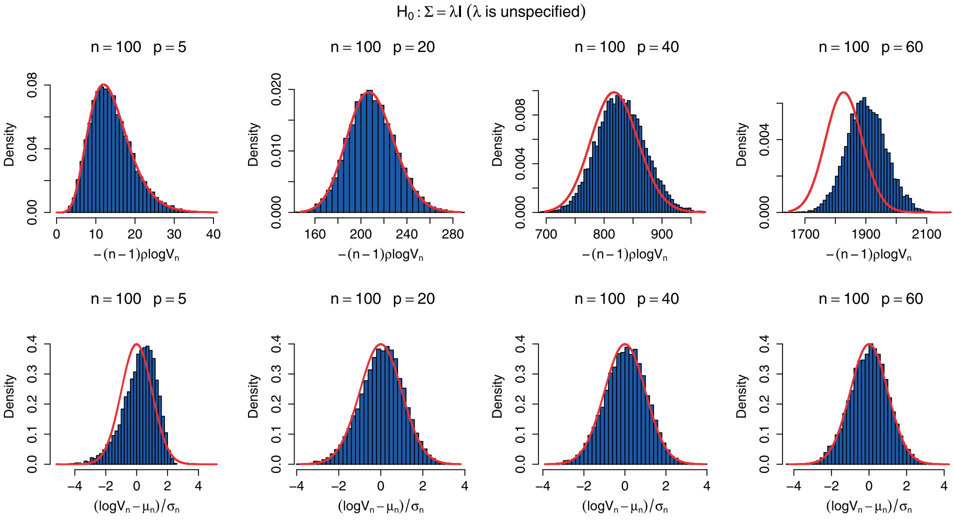


*Fig. 1.* Comparison between theorems 1 and (2.3). We choose $n = 100$ with $p = 5, 20, 40$ and 60. The pictures in the top row show that the $\chi^2$ curves stay away farther gradually from the histogram of $-(n-1)\rho \log V_n$ when $p$ grows. The pictures in the bottom row show that the $N(0,1)$-curve fits the histogram of $(\log V_n - \mu_n)/\sigma_n$ better as $p$ becomes larger.

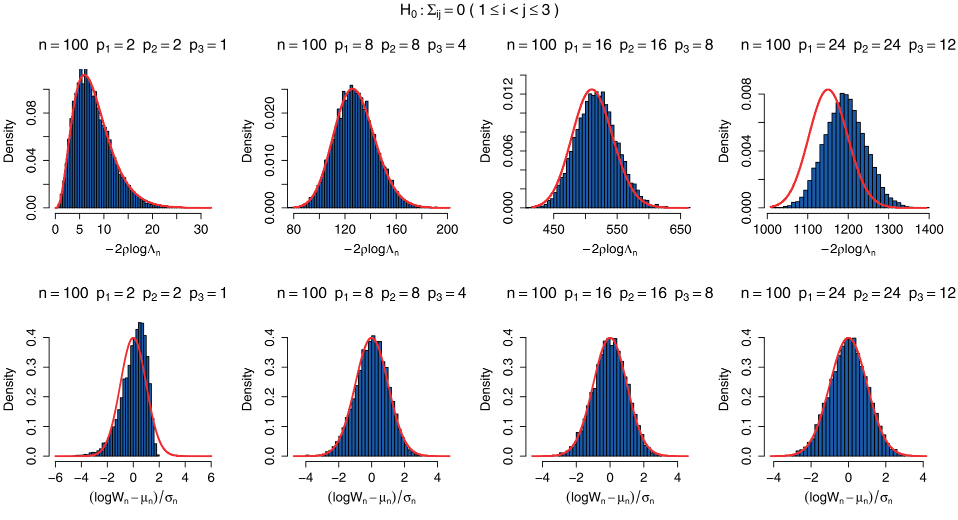


*Fig. 2.* Comparison between theorems 2 and (2.6). We choose $k = 3$, $n = 100$ and $p = 5, 20, 40$ and 60 with $p_1 : p_2 : p_3 = 2 : 2 : 1$. The pictures in the top row show that the histogram of $-2\rho \log \Lambda_n$ move away gradually from $\chi^2$ curve when $p$ grows. The pictures in the bottom row indicate that $(\log W_n - \mu_n)/\sigma_n$ and $N(0,1)$-curve fits better as $p$ becomes larger.

When $p$ is fixed, the following chi-square approximation holds under the null hypothesis of test (2.16): as $n \to \infty$,

$$-\left(n - 1 - \frac{2p+5}{6}\right) \log |\hat{\mathbf{R}}_n| \text{ converges to } \chi^2_{p(p-1)/2} \tag{2.20}$$

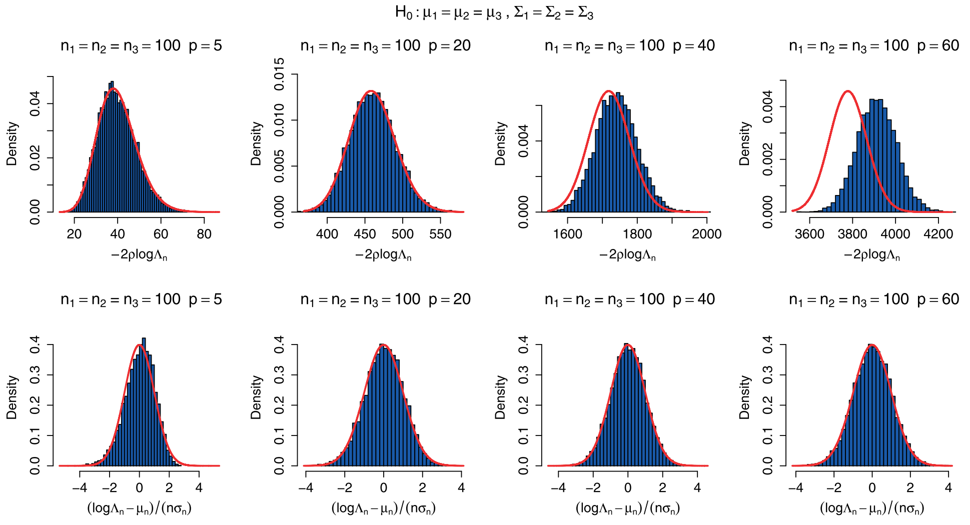


*Fig. 3.* Comparison between theorems 3 and (2.9). We choose $n_1 = n_2 = n_3 = 100$ with $p = 5, 20, 40$ and 60. The pictures in the top row show that the $\chi^2$ curves stay away farther gradually from the histogram of $-2\rho \log \Lambda_n$ when $p$ grows. The pictures in the bottom row show that the $N(0, 1)$-curve fits the histogram of $(\log \Lambda_n - \mu_n)/(n\sigma_n)$ very well as $p$ grows.

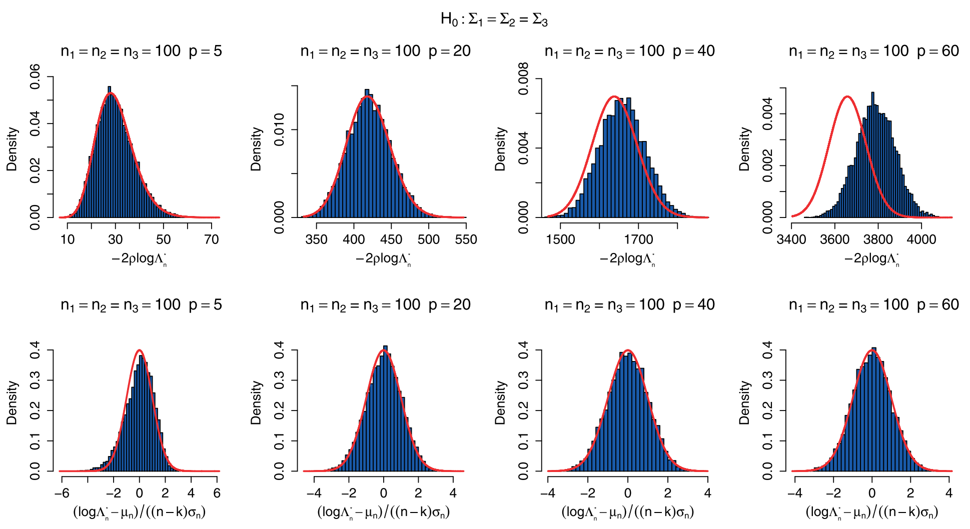


*Fig. 4.* Comparison between theorems 4 and (2.12). We chose $n_1 = n_2 = n_3 = 100$ with $p = 5, 20, 40$ and 60. The pictures in the top row show that the $\chi^2$ curves goes away quickly from the histogram of $-2\rho \log \Lambda_n^*$ as $p$ grows. The pictures in the second row show that the $N(0, 1)$-curve fits the histogram of $\left(\log \Lambda_n^* - \mu_n\right) / [(n-k)\sigma_n]$ better as $p$ grows.

in distribution. See, for example, the work of Bartlett (1954) or p. 40 from the work of Morrison (2005) for this.

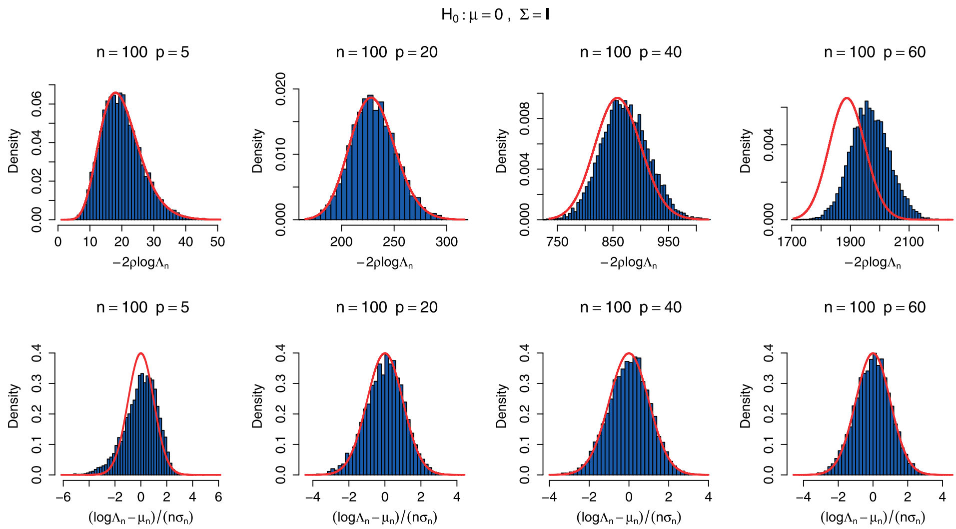


*Fig. 5.* Comparison between theorems 5 and (2.15). We choose $n = 100$ with $p = 5, 20, 40$ and 60. The pictures in the first row show that, as $p$ is large, the $\chi^2$-curve fits the histogram of $-2\rho \log \Lambda_n$ poorly. Those in the second row indicate that the N(0,1)-curve fits the histogram of $(\log \Lambda_n - \mu_n)/(n\sigma_n)$ very well as $p$ is large.

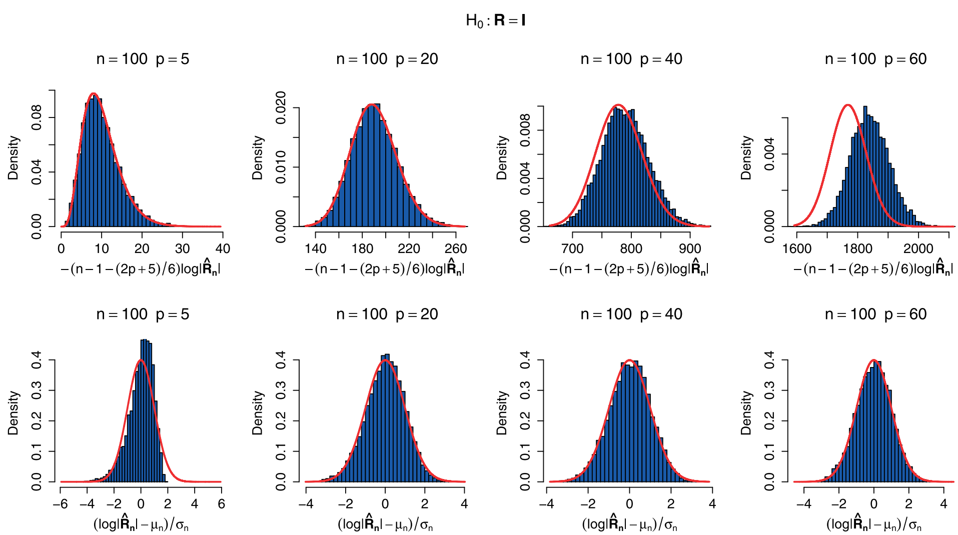


*Fig. 6.* Comparison between corollary 1 and (2.20). We choose $n = 100$ with $p = 5, 20, 40$ and 60. The pictures in the first row show that, as $p$ is large, the $\chi^2$-curve fits the histogram of $-\left(n - 1 - \frac{2p+5}{6}\right) \log |\hat{\mathbf{R}}_n|$ poorly. Those in the second row indicate that the N(0,1)-curve fits the histogram of $(\log |\hat{\mathbf{R}}_n| - \mu_n)/\sigma_n$ very well as $p$ is large.

Based on the chi-square approximation in (2.20), the LRT rejects the null hypothesis in (2.16) if $-\left(n - 1 - \frac{2p+5}{6}\right) \log |\hat{\mathbf{R}}_n| \geq \chi^2_{p(p-1)/2,\alpha}$. According to corollary 1, the rejection region based on the normal approximation is $(\log |\hat{\mathbf{R}}_n| - \mu_n)/\sigma_n \leq -z_\alpha$.

## 3. Simulation study

In this section, we compare the performance of the chi-square approximation and the normal approximation for all six LRTs in sections 2.1–2.6 through a finite sample simulation study. We plot the histograms for the six chi-square statistics that are used for the chi-square approximations specified in (2.3), (2.6), (2.9), (2.12), (2.15) and (2.20), and compare with their corresponding limiting chi-square curves. Similarly, we plot the histograms of the six statistics that are used for the normal approximations given in theorems 1–6 and compare with the standard normal curve. We also report estimated sizes and powers for the six LRTs based on their chi-square approximations and the normal approximations. All simulations have been done by using software **R** (R Development Core Team, University of Auckland, New Zealand), and the histograms, estimates of the sizes and powers, are based on 10,000 replicates.

### *3.1. Comparison of histograms*

Six figures, figures 1–6, are reported, and each of them corresponds to the two approximation methods stated in one of the sections 2.1–2.6. These figures are self-evident: when the sample sizes are small, the classical chi-square approximations are good. Our CLTs outperform the chi-square approximations when the sample sizes are large. Even the data dimensions are large but are small relative to the sample sizes, the fits are still quite well. These simulations are consistent with theorems 1–6.

### *3.2. Simulation study: Sizes and powers*

In this part, for each of the six LRTs treated earlier, we simulate the sizes and the powers for the normal approximation and for the classical chi-square approximation; the simulation results are listed in six tables, tables 1–6, which are self-explained. The notation $\mathbf{J}_p$ stands for the $p \times p$ matrix whose entries are all equal to 1, and $\lfloor x \rfloor$ stands for the integer part of $x > 0$.

Table 1. *Size and power of LRT for sphericity in section 2.1*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n = 100, p = 5$ | 0.0629 | 0.0548 | 0.7519 | 0.7340 |
| $n = 100, p = 20$ | 0.0557 | 0.0546 | 0.8757 | 0.8735 |
| $n = 100, p = 40$ | 0.0529 | 0.0868 | 0.8529 | 0.9022 |
| $n = 100, p = 60$ | 0.0570 | 0.3342 | 0.7887 | 0.9750 |

*The sizes (alpha errors) are estimated based on 10,000 simulations from $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers are estimated under the alternative hypothesis that $\mathbf{\Sigma} = diag(1.69, \cdots, 1.69, 1, \cdots, 1)$, where the number of 1.69 appearing in the diagonal is equal to $\lfloor p/2 \rfloor$.*

Table 2. *Size and power of LRT for independence of three components in section 2.2*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n = 100, p_1 = p_2 = 2, p_3 = 1$ | 0.0659 | 0.0510 | 0.7611 | 0.7199 |
| $n = 100, p_1 = p_2 = 8, p_3 = 4$ | 0.0570 | 0.0516 | 0.9787 | 0.9767 |
| $n = 100, p_1 = p_2 = 16, p_3 = 8$ | 0.0508 | 0.0699 | 0.9590 | 0.9730 |
| $n = 100, p_1 = p_2 = 24, p_3 = 12$ | 0.0539 | 0.2204 | 0.8593 | 0.9714 |

*The sizes (alpha errors) are estimated based on 10,000 simulations from $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers are estimated under the alternative hypothesis that $\mathbf{\Sigma} = \mathbf{0.15J}_p + \mathbf{0.85I}_p$.*

Table 3. *Size and power of LRT for equality of three distributions in section 2.3*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n_1 = n_2 = n_3 = 100, p = 5$ | 0.0567 | 0.0476 | 0.5857 | 0.5499 |
| $n_1 = n_2 = n_3 = 100, p = 20$ | 0.0493 | 0.0494 | 0.7448 | 0.7455 |
| $n_1 = n_2 = n_3 = 100, p = 40$ | 0.0519 | 0.0997 | 0.6645 | 0.7751 |
| $n_1 = n_2 = n_3 = 100, p = 60$ | 0.0495 | 0.4491 | 0.5134 | 0.9400 |

*The sizes (alpha errors) are estimated based on 10,000 simulations from three normal distributions of $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers were estimated under the alternative hypothesis that $\boldsymbol{\mu}_1 = (0, \ldots, 0)'$, $\boldsymbol{\Sigma}_1 = 0.5\mathbf{J}_p + 0.5\mathbf{I}_p$; $\boldsymbol{\mu}_2 = (0.1, \ldots, 0.1)'$, $\boldsymbol{\Sigma}_2 = 0.6\mathbf{J}_p + 0.4\mathbf{I}_p$; $\boldsymbol{\mu}_3 = (0.1, \ldots, 0.1)'$, $\boldsymbol{\Sigma}_3 = 0.5\mathbf{J}_p + 0.31\mathbf{I}_p$.*

Table 4. *Size and power of LRT for equality of three covariance matrices in section 2.4*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n_1 = n_2 = n_3 = 100, p = 5$ | 0.0814 | 0.0540 | 0.7239 | 0.6551 |
| $n_1 = n_2 = n_3 = 100, p = 20$ | 0.0565 | 0.0531 | 0.7247 | 0.7808 |
| $n_1 = n_2 = n_3 = 100, p = 40$ | 0.0554 | 0.0984 | 0.6111 | 0.7161 |
| $n_1 = n_2 = n_3 = 100, p = 60$ | 0.0526 | 0.4366 | 0.4649 | 0.9126 |

*The sizes (alpha errors) are estimated based on 10,000 simulations from $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers are estimated under the alternative hypothesis that $\boldsymbol{\Sigma}_1 = \mathbf{I}_p, \boldsymbol{\Sigma}_2 = 1.21\mathbf{I}_p$, and $\boldsymbol{\Sigma}_3 = 0.81\mathbf{I}_p$.*

Table 5. *Size and power of LRT for specified normal distribution in section 2.5*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n = 100, p = 5$ | 0.1086 | 0.0525 | 0.5318 | 0.3872 |
| $n = 100, p = 20$ | 0.0664 | 0.0542 | 0.7684 | 0.7375 |
| $n = 100, p = 40$ | 0.0593 | 0.0905 | 0.7737 | 0.8297 |
| $n = 100, p = 60$ | 0.0610 | 0.3478 | 0.7112 | 0.9552 |

*Sizes (alpha errors) are estimated based on 10,000 simulations from $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers are estimated under the alternative hypothesis that $\boldsymbol{\mu} = (0.1, \ldots, 0.1, 0, \ldots, 0)'$, where the number of 0.1 is equal to $\lfloor p/2 \rfloor$ and $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ where $\sigma_{ij} = 1$ for $i = j, \sigma_{ij} = 0.1$ for $0 < |i - j| \leq 3$, and $\sigma_{ij} = 0$ for $|i - j| > 3$.*

Table 6. *Size and power of LRT for complete independence in section 2.6*

| | Size under $H_0$ | | Power under $H_a$ | |
|---|---|---|---|---|
| | CLT | $\chi^2$ approx. | CLT | $\chi^2$ approx. |
| $n = 100, p = 5$ | 0.0581 | 0.0550 | 0.5318 | 0.3872 |
| $n = 100, p = 20$ | 0.0552 | 0.0558 | 0.7684 | 0.7375 |
| $n = 100, p = 40$ | 0.0512 | 0.0870 | 0.7737 | 0.8297 |
| $n = 100, p = 60$ | 0.0555 | 0.3163 | 0.7112 | 0.9552 |

*Sizes (alpha errors) are estimated based on 10,000 simulations from $N_p(\mathbf{0}, \mathbf{I}_p)$. The powers are estimated under the alternative hypothesis that the correlation matrix $\mathbf{R} = (r_{ij})_{p\times p}$, where $r_{ij} = 1$ for $i = j, r_{ij} = 0.1$ for $0 < |i - j| \leq 3$, and $r_{ij} = 0$ for $|i - j| > 3$.*

From the six tables, we can see that, when $p$ is small, the chi-square approximation works better. A common feature is that, for very small values of $p$, the LRTs based on the normal approximation have a slightly larger sizes than the nominal level 0.05. With the increase of $p$, the sizes of the LRTs based on the normal approximation method are very close to 0.05, while the sizes of the chi-square approximations are significantly higher than 0.05. We see that the sizes for the normal approximation are quite stable over the different choices of $p$.

For comparison of the powers reported in the six tables, we note that, the larger the estimated sizes, the larger the corresponding estimated powers. When $p$ is relatively large, the powers for the chi-square approximation are larger than those for the normal approximation; however, the sizes from the chi-square approximation are seriously higher than the nominal level 0.05. To understand this phenomenon, one should be aware of the fact that the two approximation methods use the same test statistics, but they result in different cutoff values for rejection regions. For illustrating purpose, we can look at the spherical test in section 2.1 with the LRT statistic $V_n$ defined in (2.2). From the last paragraph in section 2.1, the rejection region is $V_n \le c_\alpha$, where $\alpha \in (0, 1)$ is the type I error. Because $c_\alpha$ is unknown, the actual cutoffs used to approximate $c_\alpha$ are $c_{\alpha,1} = \exp\left(-\chi^2_{f,\alpha}/(n-1)\rho\right)$ from the chi-square approximation and $c_{\alpha,2} = \exp(\mu_n - z_\alpha\sigma_n)$ from the normal approximation, which result in two different rejection regions $\{V_n \le c_{\alpha,1}\}$ and $\{V_n \le c_{\alpha,2}\}$, respectively. The two rejection regions are nested, that is, one is a subset of the other. Therefore, the larger rejection region has a larger size and a larger power. In other words, the larger powers for the chi-square approximation come from the sacrifice of the accuracy in type I errors or sizes of the test. The same relation holds true for other five LRTs. This explains what we have observed in the six tables for the powers.

In what follows, we provide more explanations on the simulation results for the sizes in the six tables:

**(i)** table 1. We consider the spherical test $H_0 : \mathbf{\Sigma} = \lambda\mathbf{I}_p$ with unspecified $\lambda$, as given in (2.1). With a fixed sample size $n = 100$, we choose $p = 5, 20, 40$ and 60 for the values of the dimension $p$. When $p$ is small (5 and 20), the chi-square and the normal approximation methods are comparable, but the chi-square approximation is slightly better than the normal approximation in terms of the accuracy in the size of the test. When $p = 60$, the size for the chi-square approximation is 0.3342, much larger than the nominal level 0.05, while the size for the normal approximation is 0.0570.

**(ii)** table 2. This table reports the comparison results for the sizes and powers of the two tests for the hypotheses given in (2.4). We choose $k = 3$ for the simulation study, that is, a normal random vector is divided into three sub-vectors with dimensions $p_1, p_2$ and $p_3$, where $p_1, p_2$ and $p_3$ are specified in the table; we test the independence of the three sub-vectors. For small $p_i$'s, the chi-square approximation performs better than the normal approximation, but the normal approximation seems to be not too bad with size 0.0659 compared with the nominal level 0.05 even when $p_i$'s are as small as $p_1 = 2, p_2 = 2$ and $p_3 = 1$. The normal approximation improves as $p_i$'s grow, and, eventually, the chi-square approximation yields a much larger size than the nominal level.

**(iii)** table 3. This table is concerning the simulation on testing the equality of $k$ $p$-dimensional normal distributions as given in (2.7). We consider $k = 3$ as normal populations, and sample sizes are fixed as $n_1 = n_2 = n_3 = 100$. Then four cases, when $p = 5, 20, 40$ and 60, are investigated in the simulation; the distribution under null hypothesis used in the simulation is $N_p(\mathbf{0}, \mathbf{I}_p)$. Although the size at $p = 5$ is 0.0567, a little bit larger than the nominal level, the size for the normal approximation is quite stable in general. The size of the test based on the chi-square approximation is reasonably close to the nominal level only for very small $p$.

**(iv)** table 4. For hypotheses in (2.10), that is, testing the equality of $k$ covariance matrices of normal random vectors, we report the sizes and powers of the two tests. We choose $k = 3$ and fix the sizes of the samples for the three populations as $n_1 = n_2 = n_3 = 100$ with four different dimension choices, $p = 5, 20, 40$, and 60. The table shows that the sizes for the normal approximation are very close to the nominal level 0.05 except in the case when $p = 5$. The chi-square approximation works well only for $p = 5$, and its size grows drastically fast as $p$ increases. For example, the size for the chi-square approximation at $p = 60$ is as high as 0.9126, that is, about 91% of time, the test rejects the true null hypotheses.

**(v)** table 5. Considering the test of hypothesis that the underlying distribution is a specific multivariate normal distribution or equivalently the test described in (2.13), the table compares the performance of the two different approximation methods. We fix the sample size $n = 100$ and consider four different values for the dimension, $p = 5, 20, 40$ and 60. For small $p$, the chi-square approximation works pretty well in terms of accuracy in the size, but it becomes worse as $p$ becomes larger. The normal approximation works very well for all reasonably large $p$.

**(vi)** table 6. This table is about the test of independence of all components from a normal random vector, see (2.16) or equivalently, the covariance matrix is diagonal. In the simulation, the sample size is chosen as $n = 100$, and the dimension $p$ has four choices, $p = 5, 20, 40$ and 60. From the table, all four sizes from the normal approximation are close to 0.05, while the chi-square approximation results in a reasonable size only for $p = 5$ or 20.

## 4. Conclusions and discussions

We study six LRTs in this paper. The CLTs of the six statistics are derived under the assumption that the population dimension $p \to \infty$ and $p < n-c$ for some constant $c$ with $1 \le c \le 4$. Jiang & Yang (2013) show that the CLTs hold only at $p/n \to y \in (0, 1]$. In this paper, we obtain the same CLTs under almost to the most relaxed conditions. Precisely, if $p$ is finite, the LRT statistics converge weakly to chi-square distributions; if $p > n - 1$, the LRT statistics do no exist. The only 'sacrifice' is that $p$ is not allowed to be too close to $n$ such as $p = n - 1$ in some cases. Ignoring these small technical losses, we obtain the sufficient and necessary conditions for the CLTs. These give us almost the maximum flexibilities to use them in practice.

The strategies of our proofs are based on the moment generating functions of targeted LRT statistics. We develop a new tool in proposition 5.1, which is the key part in the proofs. This very technically involved tool is different from those used in the works of Bai *et al.* (2009), Jiang *et al.* (2012) and Jiang & Yang (2013).

Finally, we make some comments as follows:

(i) Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be iid random vectors with a probability density function or probability mass function $p(x|\theta)$, where $\theta \in \Theta$. Consider the hypothesis test $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \in \Theta \backslash \Theta_0$. Let $\Lambda_n$ be the LRT statistic. The Wilks theorem (Wilks (1938) or van der Vaart (1998)) says that $-2 \log \Lambda_n \to \chi^2_d$ as $n \to \infty$ under $H_0$, where $d$, the difference between the dimensions of $\Theta$ and $\Theta_0$, is fixed. If $d = d_n$ depends on $n$ and goes to infinity but at a very slow rate, we can see that the distributions of $-2 \log \Lambda_n$ and $\chi^2_d$ are very close. Meanwhile, by the standard CLT, $\left(\chi^2_d - d\right)/\sqrt{2d} \to N(0, 1)$

whenever $d \to \infty$. So, heuristically, $(2\log\Lambda_n + d)/\sqrt{2d} \to N(0,1)$ when $d = d_n$ goes to infinity very slowly as $n \to \infty$, it may not be true when $d$ goes to infinity too fast. This process is similar to the exchange of the limits in the real and complex analysis, which is usually non-trivial. The six theorems we present in the paper can help clarify the situation when $d$ goes to infinity, that is, even though the CLTs hold for $-2\log\Lambda_n$, the asymptotic mean and variance may not be $d$ and $2d$ anymore. The proofs of our theorems 1–6 are based on the analysis of the moments of $\Lambda_n$, which are available thanks to the normal assumptions.

(ii) In theorems 2, 3 and 4, the assumption '$\delta < p_i/p_j < \delta^{-1}$' or '$\delta < n_i/n_j < \delta^{-1}$' says that the population distribution dimensions or the sample sizes are comparable. We impose these assumptions in the theorems only for the purpose to simplify the proofs. It is possible that the conditions can be relaxed. And it will be interesting to see how less constrained among the $p_i$'s or $n_i$'s to make the three theorems hold.

(iii) Recently, some authors study similar problems under the nonparametric setting, for example, the works of Cai *et al.* (2013), Cai & Ma (2013), Chen *et al.* (2010), Li & Chen (2012), Qiu & Chen (2012) and Xiao & Wu (2013).

(iv) When the normality assumptions are slightly altered, say, the population distribution is elliptical, we expect our results still hold. If the moments of the LRT statistics are explicit, we can use a similar tool to proposition 5.1 to prove the CLTs. Otherwise, the random matrix theory can possibly be applied to obtain the CLTs, for instance, the work of Bai *et al.* (2009). In a general setting, such as the one described in item (i) earlier, investigating these LRTs can be technically complicated when the normality assumption is invalid.

(v) The CLTs in this paper are derived under null hypothesis. For tests in sections 2.5 and 2.6, we expect CLTs through similar technical tools in this paper. For the rest of the tests, the CLTs are related to the solutions of systems of partial differential equations. We leave them as a future work.

(vi) Here, we consider the scenarios when $p$ is smaller than $n$ such that either $p$ is at the same scale of $n$ or $p$ is much smaller than $n$. These are the necessary situations to study the LRTs because the tests do not exist otherwise. In a bigger picture, experts consider tests with the dimensionality of data $p$ being larger or much larger than the sample size $n$. Readers are referred to the papers, for example, by Ledoit & Wolf (2002) and Chen *et al.* (2010) for the sphericity test; by Schott (2001, 2007) for testing the equality of multiple covariance matrices and by Srivastava (2005) for testing the covariance matrix of a normal distribution. A projection method is used to investigate the two-sample test by Lopes *et al.* (2011).

(vii) Onatski *et al.* (2013) provide eigenvalue-based LRTs for the spherical test, and their asymptotic powers are studied. Another paper by Onatski *et al.* (2014) extend the results in their previous paper to some non-Gaussian cases. Further, Cai & Ma (2013) propose some relevant tests and study their optimal properties. Interested readers are referred to their papers directly.

## 5. Proofs

We only prove theorem 6. The other five theorems are proved in the Supplement by Jiang & Qi (2015). The main tool is stated next. Its proof can be seen in the Supplement. The following are some standard notation:

For two sequences of numbers $\{a_n;\, n \geq 1\}$ and $\{b_n;\, n \geq 1\}$, the notation $a_n = O(b_n)$ as $n \to \infty$ means that $\limsup_{n\to\infty} |a_n/b_n| < \infty$. The notation $a_n = o(b_n)$ as $n \to \infty$ means that $\lim_{n\to\infty} a_n/b_n = 0$, and the symbol $a_n \sim b_n$ stands for $\lim_{n\to\infty} a_n/b_n = 1$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = O(g(x))$, $f(x) = o(g(x))$ and $f(x) \sim g(x)$ as $x \to x_0 \in [-\infty, \infty]$ are similarly interpreted.

Throughout the paper $\Gamma(z)$ stands for the gamma function defined on the complex plane $\mathbb{C}$. Define the following:

$$\Gamma_p(z) := \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(z - \frac{1}{2}(i-1)\right) \tag{5.1}$$

for complex number $z$ with $\mathrm{Re}(z) > \frac{1}{2}(p-1)$. (Muirhead (1982) p. 62). The key tool to prove the main theorems is the following analysis of $\Gamma_p(z)$:

**Proposition 5.1.** *Let $\{p = p_n \in \mathbb{N};\, n \geq 1\}, \{m = m_n \in \mathbb{N};\, n \geq 1\}$ and $\{t_n \in \mathbb{R};\, n \geq 1\}$ satisfy that (i) $p_n \to \infty$ and $p_n = o(n)$; (ii) there exists $\epsilon \in (0,1)$ such that $\epsilon \leq m_n/n \leq \epsilon^{-1}$ for large $n$; and (iii) $t = t_n = O(n/p)$. Then, as $n \to \infty$,*

$$\log \frac{\Gamma_p\left(\frac{m-1}{2} + t\right)}{\Gamma_p\left(\frac{m-1}{2}\right)} = \alpha_n t + \beta_n t^2 + \gamma_n(t) + o(1),$$

*where*

$$\alpha_n = -\left[2p + \left(m - p - \frac{3}{2}\right)\log\left(1 - \frac{p}{m-1}\right)\right];\ \ \beta_n = -\left[\frac{p}{m-1} + \log\left(1 - \frac{p}{m-1}\right)\right];$$
$$\gamma_n(t) = p\left[\left(\frac{m-1}{2} + t\right)\log\left(\frac{m-1}{2} + t\right) - \frac{m-1}{2}\log\frac{m-1}{2}\right].$$

We also need the following auxiliary result: It is proved in Lemma A.1 from Supplement by Jiang & Qi.

**Lemma 5.1.** *As $x \to +\infty$,*

$$\log\frac{\Gamma(x+b)}{\Gamma(x)} = (x+b)\log(x+b) - x\log x - b - \frac{b}{2x} + O\left(\frac{b^2+1}{x^2}\right) \tag{5.2}$$

*holds uniformly on $b \in [-\delta x, \delta x]$ for any given $\delta \in (0,1)$.*

A key element in the proof of theorem 6 is the explicit expression of the moments of the determinant of the sample correlation matrix.

**Lemma 5.2.** *Let $\hat{\mathbf{R}}_n$ be the sample correlation matrix with the density function as in (2.19). Assume $n - 4 > p \geq 2$. Then,*

$$E\left[|\hat{\mathbf{R}}_n|^t\right] = \left[\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2} + t\right)}\right]^p \cdot \frac{\Gamma_p\left(\frac{n-1}{2} + t\right)}{\Gamma_p\left(\frac{n-1}{2}\right)} \tag{5.3}$$

*for all $t \geq -\max\left\{1, \left\lfloor\frac{n-p}{2}\right\rfloor - 2\right\}$.*

This lemma is a refinement of lemma 2.10 from the work of Jiang & Yang (2013) who show that (5.3) holds for all $t \geq -1$.

*Proof.* By lemma 2.10 from the work of Jiang & Yang (2013), (5.3) is true for $t \geq -1$. Thus, we only need to prove that (5.3) holds for $t \geq m := -\left\lfloor \frac{n-p}{2} \right\rfloor + 2$. It is a key observation that $m \leq 0$ by the assumption $n - 4 > p$.

Recalling (2.18), $\hat{\mathbf{R}}_n$ is a $p \times p$ non-negative definite matrix, and each of its entries takes value in $[-1, 1]$; thus, the determinant $|\hat{\mathbf{R}}_n| \leq p!$. Second, from (2.19), we see that the density function of $|\hat{\mathbf{R}}_n|$ exists; hence, $P(|\hat{\mathbf{R}}_n| = 0) = 0$. By (9) on p. 150 from the work of Muirhead (1982) or (48) on p. 492 from the work of Wilks (1932),

$$E\left[|\hat{\mathbf{R}}_n|^k\right] = \left[\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}+k\right)}\right]^p \cdot \frac{\Gamma_p\left(\frac{1}{2}(n-1)+k\right)}{\Gamma_p\left(\frac{1}{2}(n-1)\right)} \tag{5.4}$$

for any integer $k$ such that $\frac{n-1}{2} + k > \frac{p-1}{2}$ by (5.1), which is equivalent to that $k > -(n-p)/2$. Thus, (5.4) holds for all $k \geq m - 1$. In particular, $E\left[|\hat{\mathbf{R}}_n|^{m-1}\right] < \infty$. Now set $U = -\log(|\hat{\mathbf{R}}_n|/p!)$. Then $U \geq 0$ *a.s.* and $Ee^{(1-m)U} < \infty$. Because $|e^{-(z+m)U}| = e^{-(\mathrm{Re}(z)+m)U}$ and $|Ue^{-(z+m)U}| = Ue^{-(\mathrm{Re}(z)+m)U}$, they imply that

$$Ee^{-(z+m)U} \quad \text{and} \quad E\left(Ue^{-(z+m)U}\right) \quad \text{are both finite}$$

for all $\mathrm{Re}(z) \geq 0$, where we use the inequalities $e^{-mu} \leq e^{(1-m)u}$ and $ue^{-mu} \leq e^{(1-m)u}$ for all $u \geq 0$ to obtain the second assertion. Define the following:

$$h_1(z) := (p!)^{-(z+m)} \cdot E\left[|\hat{\mathbf{R}}_n|^{z+m}\right] = Ee^{-(z+m)U}$$

for all $z$ with $\mathrm{Re}(z) \geq 0$. It is not difficult to check that $\frac{d}{dz}\left(Ee^{-(z+m)U}\right) = -E\left[Ue^{-(z+m)U}\right]$ for all $\mathrm{Re}(z) \geq 0$. Furthermore, $\sup_{\mathrm{Re}(z)\geq 0} |h_1(z)| \leq Ee^{-mU} < \infty$. Therefore, $h_1(z)$ is analytic and bounded on $\{z \in \mathbb{C};\ \mathrm{Re}(z) \geq 0\}$. Define the following:

$$h_2(z) = (p!)^{-(z+m)} \cdot \left[\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}+z+m\right)}\right]^p \cdot \frac{\Gamma_p\left(\frac{n-1}{2}+z+m\right)}{\Gamma_p\left(\frac{1}{2}(n-1)\right)}$$

for $\mathrm{Re}(z) \geq 0$. By the Carlson uniqueness theorem (theorem 2.8.1 on p. 110 from the work of Andrews *et al.* (1999)), if we know that $h_2(z)$ is also bounded and analytic on $\{z \in \mathbb{C};\ \mathrm{Re}(z) \geq 0\}$, because $h_1(z) = h_2(z)$ for all $z = 0, 1, 2, \ldots$, we obtain that $h_1(z) = h_2(z)$ for all $\mathrm{Re}(z) \geq 0$. This implies our desired conclusion. Thus, we only need to check that $h_2(z)$ is bounded and analytic on $\{z \in \mathbb{C};\ \mathrm{Re}(z) \geq 0\}$. To do so, reviewing (5.1), it suffices to show that

$$h_3(z) := \prod_{i=2}^{p} \frac{\Gamma\left(\frac{n-i}{2}+z+m\right)}{\Gamma\left(\frac{n-1}{2}+z+m\right)}$$

is bounded and analytic on $\{z \in \mathbb{C};\ \mathrm{Re}(z) \geq 0\}$. Because $2 \leq \frac{n-i}{2} + m \leq \frac{n-2}{2} + m$ for all $2 \leq i \leq p$, the two properties then follow from the fact that $h(z) := \frac{\Gamma(\alpha+z)}{\Gamma(\beta+z)}$ is bounded and analytic on $\{z \in \mathbb{C};\ \mathrm{Re}(z) \geq 0\}$ for all fixed $\beta > \alpha > 0$ by lemma 3.1 from the work of Jiang & Yang (2013). $\square$

*Proof of Theorem 6.* We need to prove the following:

$$H_n := \frac{\log|\hat{\mathbf{R}}_n| - \mu_n}{\sigma_n} \text{ converges to } N(0,1) \tag{5.5}$$

in distribution as $n \to \infty$. Equivalently, it suffices to show that, for any subsequence $\{n_k\}$, there is a further subsequence $\{n_{k_j}\}$ such that $H_{n_{k_j}}$ converges to $N(0, 1)$ in distribution as $j \to \infty$.

Now, noticing $p_n/n \in [0,1]$ for all $n$, for any subsequence $n_k$, take a further subsequence $n_{k_j}$ such that $p_{n_{k_j}}/n_{k_j} \to y \in [0,1]$. So, without a loss of generality, we only need to show (5.5) under the condition that $\lim_{n\to\infty} p_n/n = y \in [0,1]$. The case for $y \in (0,1]$ is proved by Jiang & Yang (2013). We will prove the theorem for the case $y = 0$ next.

To finish the proof, it suffices to show that

$$E\exp\left\{\frac{\log|\hat{\mathbf{R}}_n| - \mu_n}{\sigma_n}s\right\} = \exp\left(-\frac{\mu_n s}{\sigma_n}\right)\cdot E\left[|\hat{\mathbf{R}}_n|^{\frac{s}{\sigma_n}}\right] \to e^{s^2/2}$$

as $n \to \infty$ for all $s$ with $|s| \le 1$, or equivalently,

$$\log E[|\hat{\mathbf{R}}_n|^t] = \mu_n t + \frac{s^2}{2} + o(1) \tag{5.6}$$

as $n \to \infty$ for all $|s| \le 1$, where $t := \frac{s}{\sigma_n}$. It is easy to see that

$$\sigma_n \sim \frac{p}{n} \text{ and } t \sim \frac{n}{p}s$$

as $n \to \infty$. In particular, $t \ge -\max\{1, \lfloor (n-p)/2 \rfloor - 2\}$ as $n$ is large enough. By lemma 5.2,

$$\log E\left[|\hat{\mathbf{R}}_n|^t\right] = -p\,\log\left[\frac{\Gamma(\frac{n-1}{2}+t)}{\Gamma\left(\frac{n-1}{2}\right)}\right] + \log\frac{\Gamma_p\left(\frac{n-1}{2}+t\right)}{\Gamma_p\left(\frac{n-1}{2}\right)}$$

if $n$ is sufficiently large. By (5.2), we have

$$p\,\log\left[\frac{\Gamma\left(\frac{n-1}{2}+t\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right] = \gamma_n(t) - \frac{npt}{n-1} + O\left(\frac{1}{p}\right),$$

where

$$\gamma_n(t) = p\left[\left(\frac{n-1}{2}+t\right)\log\left(\frac{n-1}{2}+t\right) - \frac{n-1}{2}\log\frac{n-1}{2}\right].$$

By the fact that $t \sim \frac{n}{p}s$ and in proposition 5.1,

$$\log\frac{\Gamma_p\left(\frac{n-1}{2}+t\right)}{\Gamma_p\left(\frac{n-1}{2}\right)} = -\left[2p + \left(n-p-\frac{3}{2}\right)\log\left(1-\frac{p}{n-1}\right)\right]t + \frac{\sigma_n^2 t^2}{2} + \gamma_n(t) + o(1)$$

as $n \to \infty$. Connecting the mentioned assertions, we arrive at the following:

$$\begin{aligned}\log E\left[|\hat{\mathbf{R}}_n|^t\right] &= -\left[(n-p-\frac{3}{2})\log\left(1-\frac{p}{n-1}\right) + \frac{n-2}{n-1}p\right]t + \frac{\sigma_n^2 t^2}{2} + o(1)\\ &= \mu_n t + \frac{s^2}{2} + o(1)\end{aligned}$$

as $n \to \infty$. So we obtain (5.6) and complete the proof. □

**Acknowledgements**

We are very much thankful to Fan Yang and Lin Zhang for their help in checking the mathematical proofs. We are grateful to the two referees and the Associate Editor for their constructive comments and suggestions that have led to improvements in the paper. The research of Tiefeng Jiang was supported in part by NSF Grant DMS-1208982 and DMS-1406279. The research of Yongcheng Qi was supported in part by NSF Grant DMS-1005345.

**References**

Andrews, G. E., Askey, R. & Roy, R. (1999). *Special functions*, Cambridge University Press, New York.

Bai, Z., Jiang, D., Yao, J. & Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Stat.* **37**, 3822–3840.

Bartlett, M. S. (1937). Properties and sufficiency and statistical tests. *Proc. R. Soc. Lond. A* **160**, 268–282.

Bartlett, M. S. (1954). A note on multiplying factors for various chi-squared approximations. *J. Royal Stat. Soc., Ser. B* **16**, 296–298.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317–346.

Cai, T., Liu, W. & Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. American Statistical Association* **108**, 265–277.

Cai, T. & Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19**, 2359–2388.

Chen, S., Zhang, L. & Zhong, P. (2010). Tests for high dimensional covariance matrices. *J. Amer. Stat. Assoc.* **105**, 810–819.

Fujikoshi, Y., Ulyanov, V. V. & Shimizu, R. (2010). *Multivariate statistics: high-dimensional and large-sample approximations*, Wiley, Hoboken, New Jersey.

Jiang, D., Jiang, T. & Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Stat. Plann. Inference* **142**, 2241–2256.

Jiang, T. & Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Stat.* **41**, 2029–2074.

Jiang, T. & Qi, Y. (2015). Supplement to 'Likelihood ratio tests for high-dimensional normal distributions'. Available at: http://www.stat.umn.edu/~tjiang/papers/SJSJQ.pdf.

Ledoit, O. & Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Stat.* **30**, 1081–1102.

Li, J. & Chen, S. X. (2012). Two sample tests for high dimensional covariance matrices. *Ann. Stat.* **40**, 908–940.

Lopes, M. E., Jacob, L. J. & Wainwright, M. J. (2011). *A more powerful two-sample test in high dimensions using random projection*. Available at: http://arxiv.org/abs/1108.2401.

Mauchly, J. W. (1940). Significance test for sphericity of a normal $n$-variate distribution. *Ann. Math. Stat.* **11**, 204–209.

Morrison, D. F. (2005). *Multivariate statistical methods*, (4th edn)., Duxbury Press, Brooks/Cole Thomson Learning, Belmont, CA.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley, New York.

Onatski, A., Moreira, M. J. & Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Stat.* **41**, 1204–1231.

Onatski, A., Moreira, M. J. & Hallin, M. (2014). Signal detection in high dimension: the multispiked case. *Ann. Stat.* **42**, 225–254.

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *Ann. Stat.* **8**, 247–263.

Qiu, Y.-M. & Chen, S. X. (2012). Test for bandedness of high dimensional covariance matrices with bandwidth estimation. *Ann. Stat.* **40**, 1285–1314.

Schott, J. R. (2001). Some tests for the equality of covariance matrices. *J. Stat. Plann. Inference* **94**, 25–36.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92**, 951–956.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.* **51**, 6535–6542.

Serdobolskii, V. I. (2010). *Multivariate statistical analysis: a high-dimensional approach*, Springer, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Srivastava, M. S. (2005). Some tests concerning covariance matrix in high dimensonal data. *J. Japan Statist. Soc.* **35**, 251–272.

Sugiura, N. & Nagao, H. (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Stat.* **39**, 1686–1692.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge University Press, Cambridge, UK.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24**, 471–494.

Wilks, S. S. (1935). On the independence of $k$ sets of normally distributed statistical variables. *Econometrica* **3**, 309–326.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60–62.

Xiao, H. & Wu, W. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Process. Appl.* **123**, 2899–2920.

*Received February 2014, in final form October 2014*

Tiefeng Jiang, School of Statistics, University of Minnesota, 224 Church Street, S.E., MN 55455, USA.
E-mail: jiang040@umn.edu
Yongcheng Qi, Department of Mathematics and Statistics, University of Minnesota Duluth, MN 55812, USA.
E-mail: yqi@d.umn.edu

**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher's web site including:

Proofs of Theorems 1–5