Asymptotic Distributions for Likelihood Ratio Tests for the Equality of Covariance Matrices

Wenchuan Guo[a], **Yongcheng Qi**[b]

[a]Biometrics & Data Sciences, Bristol Myers Squibb, 3551 Lawrenceville Princeton, Lawrence Township, NJ 08543, USA.

[b]Department of Mathematics and Statistics, University of Minnesota Duluth, 1117 University Drive, Duluth, MN 55812, USA.

Abstract. Consider k independent random samples from p-dimensional multivariate normal distributions. We are interested in the limiting distribution of the log-likelihood ratio test statistics for testing for the equality of k covariance matrices. It is well known from classical multivariate statistics that the limit is a chi-square distribution when k and p are fixed integers. Jiang and Yang [13] and Jiang and Qi [12] have obtained the central limit theorem for the log-likelihood ratio test statistics when the dimensionality p goes to infinity with the sample sizes. In this paper, we derive the central limit theorem when either p or k goes to infinity. We also propose adjusted test statistics which can be well approximated by chi-squared distributions regardless of values for p and k. Furthermore, we present numerical simulation results to evaluate the performance of our adjusted test statistics and the log-likelihood ratio statistics based on classical chi-square approximation and the normal approximation.

Keywords: Likelihood ratio test, central limit theorem, multivariate normal distribution, multivariate Gamma function.

1 Introduction

High-dimensional data are increasingly common in many applications of statistics, encountered particularly in biological, meteorological, and financial studies. If we have n observations from a p-dimensional population, it is natural to set up an $n \times p$ data matrix. The analysis of such matrices includes hypothesis testing, interval estimation and modeling, and is elaborated in many classical textbooks of multivariate analysis such as Anderson [2], Muirhead [16], and Eaton [9]. In these books, the properties of these tests, estimations or models are based on the assumption that the data dimensionality p is a fixed constant or is negligible compared with the sample size n. However, due to the explosive development and improvement of computing power and data collection methods, the data dimension is often high. The traditional assumption of small p large n is no longer valid because the dimensionality p can be proportionally large compared with the sample size n. Some recent results in this situation have been found in Bai *et al.* [3], Ledoit and Wolf [14], Jiang *et al.* [11], Jiang and Yang [13], Jiang and Qi [12], Dette and Dörnemann [8], and references therein.

Given that k independent samples are observed and each of them is from a p-dimensional multivariate normal distribution, we are interested in testing the equality of k covariance matrices of multivariate normal distributions under the assumption that either $p \to \infty$, or $k \to \infty$, or both. The results in this direction can be very useful when one considers multivariate analysis of variance problems for high dimensional data that are resulted from a large number of treatments.

We denote $N_p(\mu, \Sigma)$ as the multivariate normal distribution with dimensionality p, where μ is the mean vector and Σ is the covariance matrix. Assume $k \geq 2$ is an integer. Let $\{x_{i1}, \cdots, x_{in_i}\}$, $1 \leq i \leq k$ be k independent random samples from $N_p(\mu_i, \Sigma_i)$, where the mean vectors μ_i and covariance matrices Σ_i are unknown. Consider the following hypothesis test:

$$H_0 : \Sigma_1 = \cdots = \Sigma_k \quad \text{vs} \quad H_a : H_0 \text{ is not true.} \tag{1.1}$$

Define

$$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad A_i = \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)(x_{ij} - \overline{x}_i)', \ 1 \leq i \leq k,$$

and

$$A = A_1 + \cdots + A_k, \quad \text{and} \quad n = n_1 + \cdots + n_k.$$

The likelihood ratio test for (1.1), derived by Wilks [22], is as follows:

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

The details can be found in Section 8.2.2 of Muirhead [16]. The test statistic Λ_n is non-degenerate only if every A_i is of full rank, which means that $p < n_i$ is required for all $i = 1, \ldots, k$, i.e, $p < \min_{1 \leq i \leq k} n_i$. However, the drawback of this likelihood-ratio test is that it is biased when the sample sizes $n_1, \ldots, n_k$ are not all equal, which means that the probability of rejecting H_0 in (1.1) when H_0 is true can be larger than the probability of rejecting H_0 with H_0 is false. This phenomenon was first noticed by Brown [6] with $p = 1$. In what follows, we use instead the following modified likelihood ratio test statistic Λ_n^* suggested by Bartlett [4]:

$$\Lambda_n^* = \frac{\prod_{i=1}^k (|A_i|)^{(n_i-1)/2}}{(|A|)^{(n-k)/2}} \cdot \frac{(n-k)^{(n-k)p/2}}{\prod_{i=1}^k (n_i - 1)^{(n_i-1)p/2}}. \tag{1.2}$$

Λ_n^* is derived by substituting the sample size n_i by the degree of freedom $n_i - 1$ and sum of n_i by $n - k$. It is proved by Sugiura and Nagao [21] that this modified likelihood ratio test is unbiased when $k = 2$. A generalization to general integer k is made in Perlman [17].

Assume p and k are fixed integers. Under the null hypothesis in (1.1), Box [5] proved that $-2\rho \log \Lambda_n^*$ has a chi-square limit

$$-2\rho \log \Lambda_n^* \xrightarrow{d} \chi_f^2 \tag{1.3}$$

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

$$f = \frac{1}{2}p(p+1)(k-1), \tag{1.4}$$

and

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

Note that ρ converges to 1 as $\min_{1\le i\le k} n_i \to \infty$. The classic theory for likelihood ratio test statistics states that $-2\log\Lambda_n^* \xrightarrow{d} \chi_f^2$. The correction factor ρ is introduced in (1.3) for a better chi-square approximation rate.

The likelihood ratio test of size α based on the asymptotic distribution of $-2\rho\log\Lambda_n^*$ is to reject H_0 (1.1) when the sample is falling on the region

$$R_1 = \{-2\rho \log \Lambda_n^* > c_f(\alpha)\}, \tag{1.5}$$

where $c_f(\alpha)$ is the α-level upper quantile of χ_f^2 distribution. The details can be found in Muirhead [16], p.308.

In recent studies, the assumption on the fixed dimensionality p is relaxed to change with the sample size n_i so that the ratio p/n_i converges to a constant between 0 and 1, i.e.,

$$\lim_{n\to\infty} \frac{p}{n_i} = y_i \in (0, 1). \tag{1.6}$$

For instance, using the Random Matrix Theory method, the correction to the traditional likelihood ratio test for testing the equality of two covariance matrices is given in Bai *et al.* [3] as

$$L = \frac{|A|^{n_1/2} \cdot |B|^{n_2/2}}{|n_1 A/n + n_2 B/n|^{n/2}},$$

where the two sample covariance matrices are denoted as A and B, and a central limit theorem is established for $-2\log L$ under assumption (1.6). An extension to case $y_i = 1$ is also considered in Jiang, Jiang and Yang [11]. Furthermore, for likelihood ratio test for the equality of k covariance matrices, Jiang and Yang [13] and Jiang and Qi [12] have obtained the following central limit theorem (CLT)

$$\frac{\log \Lambda_n^* - a_n}{b_n} \xrightarrow{d} N(0, 1), \tag{1.7}$$

where a_n and $b_n > 0$ are asymptotic mean and standard deviation as the functions of n and p. In their work, $k \geq 2$ is a fixed integer, the dimensionality p can be proportionally large to the sample size n_i or much smaller than n_i, that is, $p/n_i \to y_i \in [0, 1]$ and $n_i - 1 \geq p$ for $1 \leq i \leq k$.

The central limit theorem (1.7) can also be true when both k and p change with the sample sizes. Guo [10] and Dette and Dörnemann [8] establish the asymptotic normality under very restrictive conditions on both k and p. See Remark 3 in Section 2 for more details.

The likelihood ratio statistic cannot be applied to the test of the homogeneity of several covariance matrices if the dimensionality p is larger than some of the sample sizes since the corresponding sample covariance matrices are singular. To overcome this problem, Schott [19], Srivastava and Yanagihara [20] propose two test statistics based on trace of sample covariance matrices. Subsequent papers such as Li and Chen [15], Cai, Liu and Xia [7], Yang and Pan [23], and Zheng *et al.* [24] have developed various test statistics which can be also applied to non-normal populations.

Our objective in this paper is to investigate the limiting distribution of the log-likelihood ratio test statistic $-2\log\Lambda_n^*$ for the multivariate normal distributions under the minimal conditions. We have pointed out that the likelihood ratio test exists only if $1 \leq p < \min_{1\leq i\leq k} n_i$. In this paper, we assume that both p and k can change with sample sizes, $1 \leq p < \min_{1\leq i\leq k} n_i$, and $\min_{1\leq i\leq k} n_i$ tends to infinity. We will show in Theorem 2.1 that $-2\log\Lambda_n^*$, after properly normalized, converges in distribution to the standard normal distribution if $\max(p, k) \to \infty$. Without imposing this additional condition, we propose an adjusted log-likelihood ratio test statistic (ALRT) in (2.7) in Section 2 and prove in Theorem 2.2 that the ALRT statistic is approximately distributed a chi-square with $\frac{1}{2}p(p+1)(k-1)$ degrees of freedom. Neither $-2\log\Lambda_n^*$ nor $-2\rho\log\Lambda_n^*$ shares this property when p or k are too large.

The rest of the paper is organized as follows. In Section 2, we give the main results in the paper, including the CLT for $-2\log\Lambda_n^*$ and chi-square approximation to the ALRT statistics and some discussions; In Section 3, we carry out a simulation study to compare the performance of $-2\log\Lambda_n^*$ based on the classical chi-square approximation, the normal approximation and the ALRT statistics based on the chi-square approximation. All proofs are provided in Section 4.

2 Main results

We first introduce some notations. Let $\Gamma(x)$ denote the gamma function given by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt, \quad x > 0.$$

Define the digamma function ψ by

$$\psi(x) = \frac{d\log\Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}, \quad x > 0. \tag{2.1}$$

Define

$$\xi(x) = -2(\log(1-x) + x), \qquad x \in [0, 1). \tag{2.2}$$

$\xi(x)$ is nonnegative in $[0, 1)$.

Throughout this paper, we write $n = \sum_{i=1}^n n_i$ and assume that the minimal sample size $\min_{1\le i\le k} n_i$ goes to infinity. Furthermore, both the dimensionality parameter p and number of the populations k can diverge. For this reason, we write $p = p_n$ and $k = k_n$ to indicate that both parameters may change with the sample sizes.

Define

$$\begin{aligned} \mu_n &= (n-k_n)\sum_{j=1}^{p_n}\psi(\frac{n-k_n+1-j}{2}) - \sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n}\psi(\frac{n_i-j}{2}) \\ &\quad + p_n\sum_{i=1}^{k_n}(n_i-1)\log(n_i-1) - p_n(n-k_n)\log(n-k_n) \end{aligned} \tag{2.3}$$

and

$$\sigma_n^2 = \sum_{i=1}^{k_n}(n_i-1)^2\xi(\frac{p_n}{n_i}) - (n-k_n)^2\xi(\frac{p_n}{n-k_n+1}) + p_n(k_n-1). \tag{2.4}$$

Theorem 2.1. *Assume* $1 \le p_n < \min_{1\le i\le k_n} n_i$ *and* $\min_{1\le i\le k_n} n_i \to \infty$. *Let* Λ_n^* *be as in (1.2). Then, under* H_0 *in (1.1),*

$$\frac{-2\log\Lambda_n^* - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1) \qquad \text{as } n\to\infty \tag{2.5}$$

if and only if $\max(p_n, k_n) \to \infty$ *as* $n \to \infty$.

Based on Theorem 2.1, the rejection region of the test (1.1) at level α is

$$R_2 = \{(-2\log\Lambda_n^* - \mu_n)/\sigma_n > z_\alpha\}, \tag{2.6}$$

where z_α is α level critical value of the standard normal distribution.

The chi-square appropriation in (1.3) and the central limit theorem in Theorem 2.1 exhibit two different types of limiting distributions for inference. As our simulation study in Section 3 indicates, when $\max(k_n, p_n)$ is small, the chi-square approximation (1.3) works very well; while $\max(p_n, k_n)$ is large, the normal approximation prevails. In practice, one needs to decide which approximation should be used. The two theoretical results don't clearly suggest when to use the normal approximation and when to use the chi-square approximation. It is desirable if we are able to construct a test statistic which is a linear function of $-2\log\Lambda_n^*$ with a unified asymptotic distribution regardless of values of p_n and k_n. The chi-square distributions are better choices than the normal distribution for this purpose since $-2\log\Lambda_n^*$ is not asymptotically normal when both p and k are small. We note that the ratio of the mean and the variance of a chi-square distribution is $1:2$. However, neither $-2\log\Lambda_n^*$ nor $-2\rho\log\Lambda_n^*$ is well approximated by a chi-square distribution when $\max(p_n, k_n)$ is too large as the ratio of their asymptotic means and variances is not very close to $1:2$ in this case.

We propose the following adjusted log-likelihood ratio test (ALRT) statistic

$$Z_n = (-2\log\Lambda_n^*)\sqrt{\frac{2f_n}{\sigma_n^2}} + f_n - \mu_n\sqrt{\frac{2f_n}{\sigma_n^2}}, \tag{2.7}$$

where $f_n = \frac{1}{2}p_n(p_n+1)(k_n-1)$ as given in (1.4), μ_n and σ_n^2 are defined in (2.3) and (2.4), respectively. Note that the ALRT is essentially the LRT statistic since it is a linear function of the log-likelihood ratio test statistic, $-2\log\Lambda_n^*$. This new test statistic achieves an asymptotic mean $\frac{1}{2}p_n(p_n+1)(k_n-1)$ and an asymptotic variance $p_n(p_n+1)(k_n-1)$. We have the following theorem regarding the chi-square approximation of Z_n statistic.

Theorem 2.2. *Assume $p_n \ge 1$ and $k_n \ge 2$ are two sequences of integers, $p_n < \min_{1\le i\le k_n} n_i$ and $\min_{1\le i\le k_n} n_i \to \infty$. Then we have under H_0 in (1.1) that*

$$\lim_{n\to\infty} \sup_{-\infty<x<\infty} |P(Z_n \le x) - P(\chi^2_{f_n} \le x)| = 0. \tag{2.8}$$

Based on Theorem 2.2, the rejection region of the test (1.1) at level α is

$$R_3 = \{Z_n > c_{f_n}(\alpha)\}, \tag{2.9}$$

where $c_{f_n}(\alpha)$ is the α-level critical value of a chi-square distribution with $f_n = \frac{1}{2}p_n(p_n+1)(k_n-1)$ degrees of freedom.

Under some restrictive conditions, we can avoid using the digamma function when we approach the mean of $-2\log\Lambda_n^*$. We have the following theorem.

Theorem 2.3. *Assume p_n and $k_n \ge 2$ are two sequences of positive integers such that $\min_{1\le i\le k_n} n_i \to \infty$, $\sqrt{k_n}/\min_{1\le i\le k_n} n_i \to 0$ as $n\to\infty$, and there exists a constant $\delta\in(0,1)$ such that $p_n \le \delta\min_{1\le i\le k_n} n_i$ for all large n. Define*

$$\bar{\mu}_n = (n-k_n)(p_n-n+k_n+\frac{1}{2})\log(1-\frac{p_n}{n-k_n+1}) - \sum_{i=1}^{k_n}(n_i-1)(p_n-n_i+\frac{3}{2})\log(1-\frac{p_n}{n_i}) - p_n(k_n-1). \tag{2.10}$$

(a). If Z_n given in (2.7) *is redefined with μ_n being replaced by $\bar{\mu}_n$, then* (2.8) *holds under H_0 in* (1.1)*;*
(b). Additionally, if $\max(p_n, k_n) \to \infty$ *as $n\to\infty$, then* $(-2\log\Lambda_n^* - \bar{\mu}_n)/\sigma_n$ *converges to* $N(0,1)$ *under H_0 in* (1.1).

Remark 1. In both Theorems 2.1 and 2.2, we assume $1 \le p_n < \min_{1\le i\le k_n} n_i$ and $\min_{1\le i\le k_n} n_i \to \infty$ as $n\to\infty$. The first condition ensures the likelihood ratio test statistics exist and the second condition guarantees that the minimal sample size goes to infinity so that the asymptotic properties of $-2\log\Lambda_n^*$ can be investigated. Both conditions are necessary and cannot be weakened. Theorem 2.1 shows that a necessary and sufficient condition for the central limit theorems is $\lim_{n\to\infty}\max(p_n,k_n) = \infty$, which is equivalent to $\lim_{n\to\infty} f_n = \infty$, where $f_n = \frac{1}{2}p_n(p_n+1)(k_n-1)$. If f_n is bounded or fixed, the distribution of $-2\log\Lambda_n^*$ is well approximated by $\chi^2_{f_n}$ distribution. More generally, Theorem 2.2 ensures that the ALRT statistic, Z_n as defined in (2.7), is well approximately by $\chi^2_{f_n}$.

Remark 2. In both Theorems 2.1 and 2.2, the digamma function ψ as defined in (2.1) is involved in the computation of the asymptotic mean μ_n of $-2\log\Lambda_n^*$ via equation (2.3). We notice that the digamma function ψ is built in software **R** and can be calculated using the `digamma` function in **R**.

Remark 3. Assume the null hypothesis in (1.1) holds, $1 \le p_n < \min_{1\le i\le k_n} n_i$, and $\min_{1\le i\le k_n} n_i \to \infty$. Our Theorem 2.1 indicates that $\lim_{n\to\infty}\max(p_n, k_n) = \infty$ is the necessary and sufficient for the central limit limit theorems of $-2\log\Lambda_n^*$. In Guo [10], it is shown that

$$\frac{-2\log\Lambda_n^* - \bar{\mu}_n}{\sigma_n} \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty$$

under additional conditions that $k_n/\min_{1\le i\le k_n} n_i \to 0$, $\min(p_n, k_n) \to \infty$ as $n \to \infty$, and there exists a constant $\delta \in (0,1)$ such that $p_n \le \delta \min_{1\le i\le k_n} n_i$ for all large n. Theorem 3 in Dette and Dörnemann [8] states that $-2\log\Lambda_n^*$, after properly normalized, converges in distribution to the standard normal under a different set of additional assumptions that $p_n/(n-k_n) \to 0$, $\min(p_n, k_n) \to \infty$, $k_n/\sqrt{p_n(n-k_n)} \to 0$ as $n \to \infty$, and there exists a constant $\delta \in (0,1)$ such that $p_n \le \delta \min_{1\le i\le k_n} n_i$ for all large n. Clearly, either in Guo [10] or Dette and Dörnemann [8], the following situations are not under consideration:
(a) $\lim_{n\to\infty} p_n/\min_{1\le i\le k_n} n_i = 1$;
(b) $\lim_{n\to\infty} k_n/(p_n \max_{1\le i\le k_n} n_i) > 0$.

Remark 4. As we have mentioned earlier, conditions $1 \le p_n < \min_{1\le i\le k_n} n_i$ and $\min_{1\le i\le k_n} n_i \to \infty$ are necessary for discussion on the asymptotic distribution of $-2\log\Lambda_n^*$. Theorem 2.2 indicates that the new test statistic Z_n can be approximated by chi-square distribution uniformly over $1 \le p_n < \min_{1\le i\le k_n} n_i$ and $k_n \ge 2$ as $\min_{1\le i\le k_n} n_i \to \infty$. However, there is a transition from chi-square to the normal in the limiting distribution of $-2\log\Lambda_n^*$.

3 Simulation study

In this section, we carry out some simulation studies to compare the performance for the likelihood ratio test statistics based on three different approximations, including the classic chi-square approach (1.3), the CLT approach (2.5) and ALRT approach (2.8). The corresponding rejection regions for a size α test on (1.1) are given in equations (1.5), (2.6) and (2.9), respectively.

We compare the three test statistics as follows: a). We plot the histograms of the three statistics and see how well the histograms for each statistic fits its theoretical distribution under the null hypothesis of (1.1); b). We estimate the sizes (or type I errors) and determine which test statistic is more accurate; c). We estimate the power for each test statistic under some alternative hypotheses of (1.1).

For each combination of p and k with $p = 3, 20, 50, 95$ and $k = 5, 20, 50$, we set $n_i = 100$ for $1 \le i \le k$. We generate k random samples with sample sizes $n_1, \cdots, n_k$ respectively from multivariate normal distributions with selected covariance matrices, and then we calculate three test statistics. Histograms, estimation of sizes and powers are based on $10,000$ replicates. In our histograms and tables for sizes and powers for three test statistics, we use "Chisq", "CLT" and "ALRT" to denote the classic chi-square approximation (1.3), normal approximation (2.5), and adjusted chi-square approximation (2.8), respectively.

3.1 Histograms and empirical sizes

We generate random samples from k multivariate normal distributions under the null hypothesis of (1.1). Denote the common variance matrix as Σ which is a p by p matrix. Since the distribution function of Λ_n^* doesn't depend on the matrix Σ, we can simply set $\Sigma = I_p$ as a p by p identity matrix.

For each combination of p and k, we plot the histograms for the three standardized test statistics, namely $-2\rho \log \Lambda_n^*$, $(-2\log \Lambda_n^* - \mu_n)/\sigma_n$, and Z_n. The histograms for these statistics are also compared with their approximate density functions – a chi-squared distribution with $\frac{1}{2}p(p+1(k-1)$ degrees of freedom, the standard normal distribution, and a chi-squared distribution with $f = \frac{1}{2}p(p+1)(k-1)$ degrees of freedom, respectively, as established in (1.3), (2.5) and (2.8). See Figure 1 and Figures S2 and S3 in the supplement.

The histograms in Figure 1 and Figures S2 and S3 reveal the overall performance for the three test statistics. We can conclude that the classic chi-square approximation works well when p is small and gets worse as p increases, the normal approximation starts to get better as $f = \frac{1}{2}p(p+1)(k-1)$ increases, and only the histograms for the adjusted test statistic fit the chi-squared curves in all cases.

The empirical sizes for all three tests at level $\alpha = 0.05$ are reported in Table 1. A good test in terms of the type I errors (or sizes) will have an actual size close to its nominal level 0.05. The adjusted test statistic outperforms over the other two test statistics since its sizes are close to 0.05 in all cases. In particular, the classic chi-square approximation performs well only when p is relatively small. When p is large, the null hypothesis is rejected with a probability close to one.

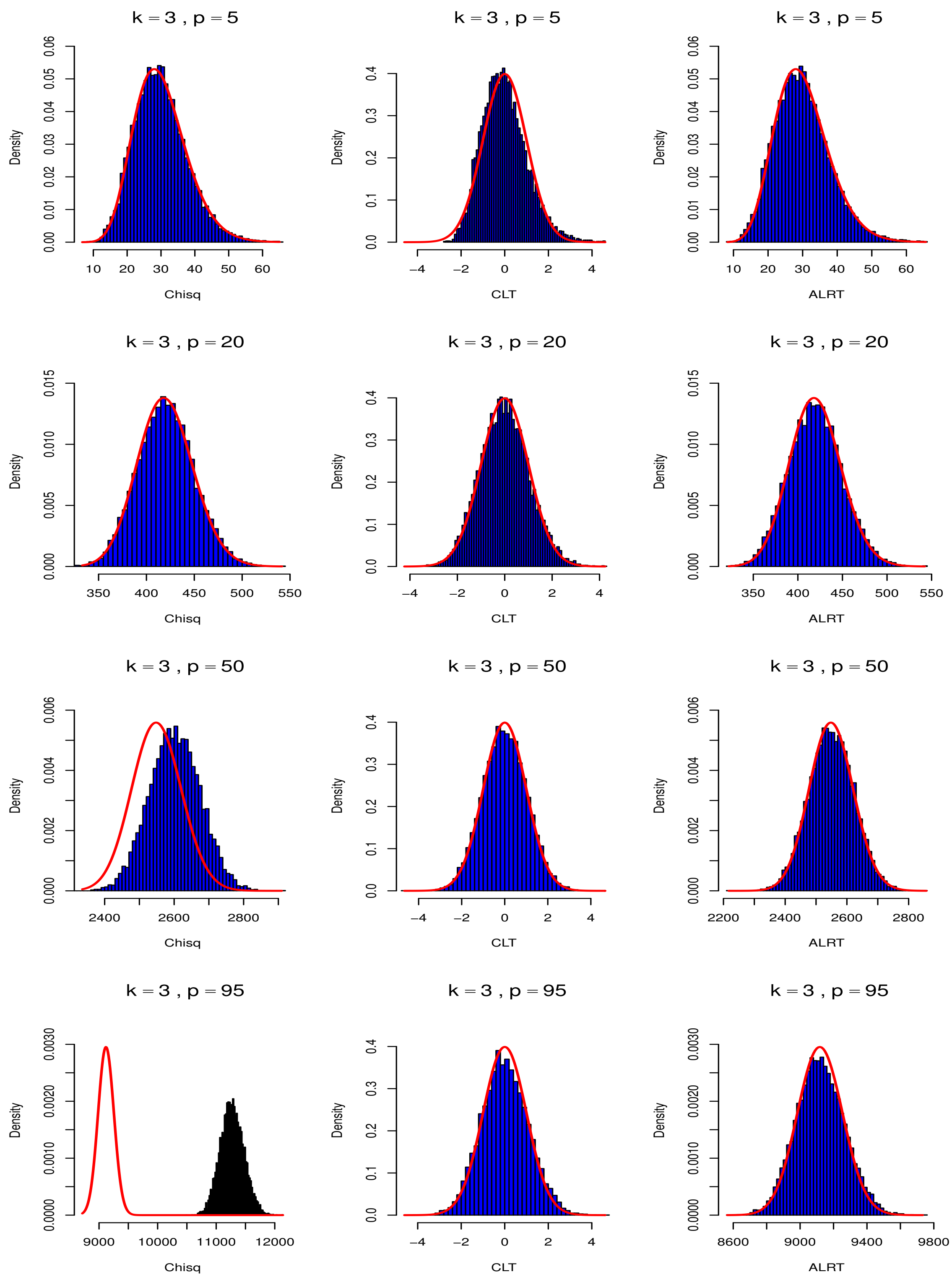

Figure 1: Histograms of $-2\rho \log \Lambda_n^*$ (Chisq), $(-2\log \Lambda_n^* - \mu_n)/\sigma_n$ (CLT), and Z_n (ALRT), where $n_i = 100$, $1 \le i \le k$, with $k = 3$ and $p = 5, 20, 50, 95$.

Table 1: The size of LRT for equality of k covariance matrices

Conditions	Chisq	CLT	ALRT
$n_i = 100,\ k = 3,\ p = 5$	0.0501	0.0636	0.0529
$n_i = 100,\ k = 3,\ p = 20$	0.0575	0.0608	0.0569
$n_i = 100,\ k = 3,\ p = 50$	0.1861	0.0539	0.0521
$n_i = 100,\ k = 3,\ p = 95$	1.0000	0.0689	0.0678
$n_i = 100,\ k = 20,\ p = 5$	0.0511	0.0576	0.0529
$n_i = 100,\ k = 20,\ p = 20$	0.0639	0.0542	0.0530
$n_i = 100,\ k = 20,\ p = 50$	0.8135	0.0533	0.0529
$n_i = 100,\ k = 20,\ p = 95$	1.0000	0.0618	0.0614
$n_i = 100,\ k = 50,\ p = 5$	0.0490	0.0540	0.0505
$n_i = 100,\ k = 50,\ p = 20$	0.0755	0.0525	0.0519
$n_i = 100,\ k = 50,\ p = 50$	0.9937	0.0554	0.0551
$n_i = 100,\ k = 50,\ p = 95$	1.0000	0.0619	0.0618

3.2 Comparison of powers

The power of a test is the probability of rejecting H_0 when H_0 is false. For each combination of k and p, we take k random samples, and the i-th sample is from a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_i = I_p + \frac{0.05(i-1)}{\sqrt{k}} J_p$ for $1 \le i \le k$, where I_p is a p by p identity matrix, and J_p is a p by p matrix with all entries being equal to 1. Again, all sample sizes are the same with $n_1 = \cdots = n_k = 100$ in the study. The powers for the three test statistics are estimated based on $10,000$ replicates, and they are reported in Table 2.

We first note that all three test statistics are linear combinations of $-2\log\Lambda_n^*$, the corresponding rejection regions at level α are of the same type as $\{-2\log\Lambda_n^* > c\}$. The only difference is that the values of cutoff c corresponding to the three approximations are different. That means the three rejections R_1, R_2 and R_3, defined in (1.5), (2.6) and (2.9), respectively, are nested each other. A large rejection region has a larger size of test and a larger power as well. From Table 2, we can see that whenever any two test statistics have really close sizes, they have similar powers. When the classic chi-square approximation fails, that is, it produces a large size, it also yields a large power. However, its large power does not make any sense here. So we compare the normal approximation and the chi-square approximation based on the test statistic Z_n. We observe that both test statistics have comparable powers when the normal approximation achieves a reasonable size.

Finally, we will demonstrate how the powers of the three test statistics change with p or k while other variables remain constant. For illustration purpose, we set $k = 10$ and $n_i = 100$ when p increases from 10 to 90, and set $p = 50$ and $n_i = 100$ when k increases from 10 to 90. The covariance structures are the same as specified in the beginning of this subsection. The curves of the powers for the three test statistics are plotted in Figure 2. We see that the power of the classic chi-square statistic (Chisq) is much larger when p is bigger, and this is the case for the two plots in the figure. Keep in mind that we have set $p = 50$ in the second plot in Figure 2. This phenomenon happens since the classic chi-square approach results in a much larger type I error than the nominal level 0.05 as in Table 1. Two other methods are very close in terms of power.

Table 2: The power of LRT for equality of k covariance matrices

Conditions	Chisq	CLT	ALRT
$n_i = 100,\ k = 3,\ p = 5$	0.0744	0.0958	0.0778
$n_i = 100,\ k = 3,\ p = 20$	0.1203	0.1260	0.1194
$n_i = 100,\ k = 3,\ p = 50$	0.3249	0.1233	0.1197
$n_i = 100,\ k = 3,\ p = 95$	1.0000	0.0987	0.0976
$n_i = 100,\ k = 20,\ p = 5$	0.5587	0.5809	0.5656
$n_i = 100,\ k = 20,\ p = 20$	0.7025	0.6728	0.6690
$n_i = 100,\ k = 20,\ p = 50$	0.9858	0.3798	0.3775
$n_i = 100,\ k = 20,\ p = 95$	1.0000	0.1478	0.1473
$n_i = 100,\ k = 50,\ p = 5$	0.9992	0.9992	0.9992
$n_i = 100,\ k = 50,\ p = 20$	0.9971	0.9951	0.9951
$n_i = 100,\ k = 50,\ p = 50$	1.0000	0.7544	0.7534
$n_i = 100,\ k = 50,\ p = 95$	1.0000	0.2284	0.2278

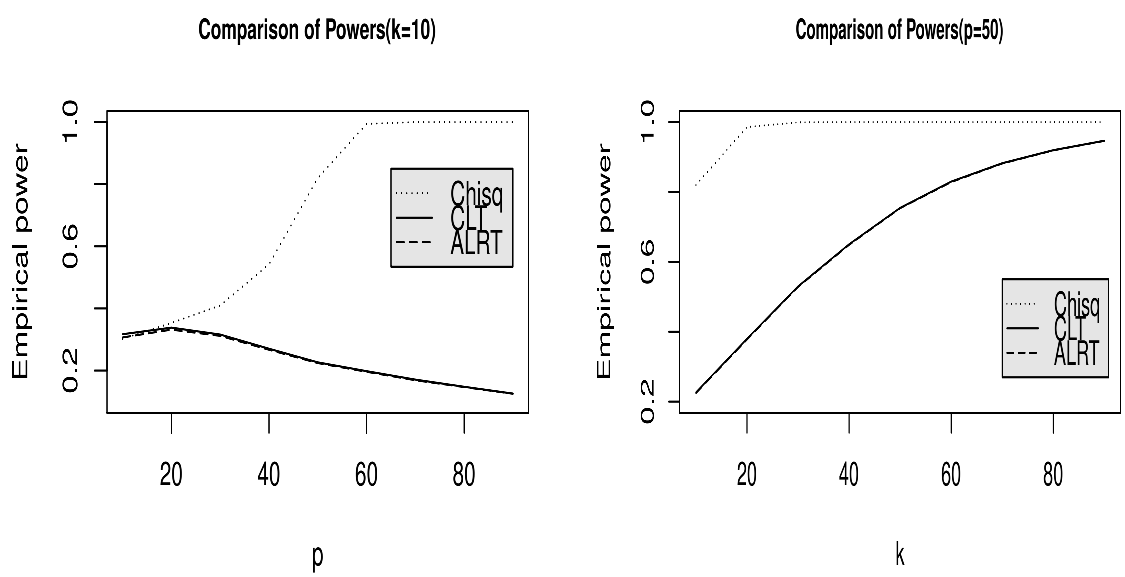

Figure 2: Plots of powers for $-2\rho \log \Lambda_n^*$ (Chisq), $(-2\log \Lambda_n^* - \mu_n)/\sigma_n$ (CLT), and Z_n (ALRT) with $n_1 = \cdots = n_k = 100$. The left one shows how the powers change with p when $k = 10$, and the right one displays how the powers change with k when $p = 50$.

3.3 Concluding remark

In this paper, we study the limiting distribution of the likelihood ratio test statistics for testing equality of the covariance matrices from k multivariate normal populations. We allow both k and the dimensionality p of the populations to change with the sample sizes. We establish the central limit theorem for the log-likelihood test when $\max(k,p)$ tends to infinity. Further, we propose an adjusted log likelihood test statistic and prove that the distribution of the new test statistic can be well approximated by a chi-squared distribution regardless of how k and p change with the sample sizes. This new test outperforms in terms of accuracy in type I error and has the comparable power as the classic chi-square approximation or the normal approximation when they are applicable.

4 Proofs

The standard notation to describe the limiting behavior of a function is denoted by big O and small o. Precisely, for two sequences of numbers $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$, $a_n = O(b_n)$ as $n \to \infty$ means that $\limsup_{n\to\infty} |a_n/b_n| < \infty$ and $a_n = o(b_n)$ as $n \to \infty$ means that $\lim_{n\to\infty} a_n/b_n = 0$. The symbol $a_n \sim b_n$ stands for $\lim_{n\to\infty} a_n/b_n = 1$. For two functions f and g, we also use $f(x) = O(g(x))$, $f(x) = o(g(x))$, and $f(x) \sim g(x)$ with the similar meaning as $x \to x_0 \in [-\infty, \infty]$.

The multivariate gamma function, denoted by $\Gamma_p(z)$, is defined as

$$\Gamma_q(z) = \pi^{q(q-1)/4} \prod_{i=1}^{q} \Gamma\Big(z - \frac{1}{2}(i-1)\Big), \tag{4.1}$$

for complex numbers z with $\mathrm{Re}(z) > \frac{1}{2}(q-1)$, where $q \geq 1$ is an integer. In our applications we only need to consider $\Gamma_q(z)$ for real variable $z > \frac{q-1}{2}$. We will put much effort in expanding the function $\Gamma_q(z)$ below.

The function $\xi(x)$ defined in (2.2) is nonnegative in $x \in [0,1)$. It is easy to see that $\xi(0) = 0$ and $\xi'(x) = \frac{2x}{1-x}$. We have

$$\xi(x) = \xi(x) - \xi(0) = \int_0^x \xi'(t)dt = 2\int_0^x \frac{t}{1-t}dt.$$

By writing $t = ux$, we have $dt = xdu$, and

$$\xi(x) = 2\int_0^1 \frac{ux^2}{1-ux}du = 2x^2\int_0^1 \frac{u}{1-ux}du, \qquad x \in [0,1).$$

Now we define

$$\eta(x) = \frac{\xi(x)}{x^2} = 2\int_0^1 \frac{u}{1-ux}du, \qquad x \in [0,1). \tag{4.2}$$

We can conclude that

$$\eta(0) = 1, \quad \lim_{x\uparrow 1} \eta(x) = \infty, \text{ and } \eta(x) \geq 1 \text{ is increasing in } [0,1). \tag{4.3}$$

4.1 Some auxiliary lemmas

Lemma 4.1. *Assume* $\{r_m,\ m \geq 2\}$ *is any sequence of integers such that* $1 \leq r_m < m$ *and* $r_m/m \to$ *as* $m \to \infty$. *Then for any* $\delta > 0$

$$\max_{r_m \leq q < m} \Big(\frac{m}{q(m-q)}\Big)^{\delta} \frac{1}{\eta(\frac{q}{m})} \to 0 \quad \text{as } m \to \infty,$$

where η *is defined in* (4.2).

Proof. One can easily verify

$$\frac{m}{q(m-q)} \leq \frac{m}{m-1} \leq 2, \qquad \text{for } 1 \leq q < m \tag{4.4}$$

and

$$\min_{r_m \leq q \leq m-r_m} \frac{m}{q(m-q)} \leq \frac{m}{r_m(m-r_m)} \to 0 \quad \text{as } m \to \infty. \tag{4.5}$$

Then it follows from equations (4.3), (4.4) and (4.5) that

$$\begin{aligned}
\max_{r_m \leq q < m} \Big(\frac{m}{q(m-q)}\Big)^{\delta} \frac{1}{\eta(\frac{q}{m})} &= \max\Big(\max_{r_m \leq q \leq m-r_m} \Big(\frac{m}{q(m-q)}\Big)^{\delta} \frac{1}{\eta(\frac{q}{m})}, \max_{m-r_m \leq q \leq m} \Big(\frac{m}{q(m-q)}\Big)^{\delta} \frac{1}{\eta(\frac{q}{m})}\Big) \\
&\leq \max\Big(\Big(\max_{r_m \leq q \leq m-r_m} \frac{m}{q(m-q)}\Big)^{\delta}, \frac{2^{\delta}}{\eta(\frac{m-r_m}{m})}\Big) \\
&\to 0
\end{aligned}$$

as $m \to \infty$. This completes the proof. ∎

Lemma 4.2. *As* $m \to \infty$

$$\max_{1 \leq q < m} \frac{q^2}{m^2(m-q)} \frac{1}{\xi(\frac{q}{m})} = o(1). \tag{4.6}$$

and

$$\max_{1 \leq q < m} \frac{q}{m(m-q)^2} \frac{1}{\xi(\frac{q}{m})} = o(1). \tag{4.7}$$

Proof. Let r_m be any sequence of integers satisfying Lemma 4.1. Since

$$\frac{q^2}{m^2(m-q)} \frac{1}{\xi(\frac{q}{m})} = \frac{1}{m-q} \frac{1}{\eta(\frac{q}{m})}$$

and $\eta(x) \geq 1$, we have

$$\begin{aligned}
\max_{1 \leq q < m} \frac{q^2}{m^2(m-q)} \frac{1}{\xi(\frac{q}{m})} &= \max\Big(\max_{r_m \leq q < m} \frac{q}{m} \frac{m}{q(m-q)} \frac{1}{\eta(\frac{q}{m})}, \max_{1 \leq q < r_m} \frac{1}{m-q} \frac{1}{\eta(\frac{q}{m})}\Big) \\
&\leq \max\Big(\max_{r_m \leq q < m} \frac{m}{q(m-q)} \frac{1}{\eta(\frac{q}{m})}, \max_{1 \leq q < r_m} \frac{1}{m-q}\Big) \\
&= \max\Big(\max_{r_m \leq q < m} \frac{m}{q(m-q)} \frac{1}{\eta(\frac{q}{m})}, \frac{1}{m-r_m+1}\Big),
\end{aligned}$$

which converges to zero as $m \to \infty$ according to Lemma 4.1. This proves (4.6).

(4.7) can be proved in the same manner and the details are omitted. ∎

Lemma 4.3. *For any $x_0 > 0$, there exist constants C_1 and C_2 such that*

$$|\psi(x) - \log x + \frac{1}{2x} + \frac{1}{12x^2}| \le \frac{C_0}{x^4}, \quad x \ge x_0; \tag{4.8}$$

$$|\psi'(x) - \frac{1}{x} - \frac{1}{2x^2}| \le \frac{C_1}{x^3}, \quad x \ge x_0. \tag{4.9}$$

Proof. Without loss of generality, assume $x_0 \in (0, 1)$.

It follows from Formulas 6.3.18 and 6.4.12 in Abramowitz and Stegun [1] that

$$\psi(x) = \log x - \frac{1}{2x} - \frac{1}{12x^2} + O(\frac{1}{x^4}),$$

$$\psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + O(\frac{1}{x^3})$$

as $x \to \infty$. Therefore, there exist constant $x_1 \ge 1$ and $C > 0$ such that (4.8) and (4.9) hold for $x \ge C$ with $C_0 = C_1 = C$. It remains to show that (4.8) and (4.9) hold for $x_0 \le x \le x_1$. We show (4.8) only for illustration. Since the function $\psi(x) - \log x + \frac{1}{2x} + \frac{1}{12x^2}$ is continuous in $[x_0, x_1]$, we have $C_3 := \sup_{x\in[x,x_1]} |\psi(x) - \log x + \frac{1}{2x} + \frac{1}{12x^2}| < \infty$. By setting $C_0 = \max(C, x_1^4 C_3)$, we have

$$|\psi(x) - \log x + \frac{1}{2x} + \frac{1}{12x^2}| \le C_3 \le \frac{C_3 x_1^4}{x^4} = \frac{C_0}{x^4}, \quad x \in [x_0, x_1].$$

Obviously, we also have

$$|\psi(x) - \log x + \frac{1}{2x} + \frac{1}{12x^2}| \le \frac{C}{x^4} \le \frac{C_0}{x^4}, \quad x \in [x_1, \infty).$$

This proves (4.8). The proof of Lemma 4.3 is done. ∎

Lemma 4.4. *Define*

$$s(x) = (x - \frac{1}{2}) \log(x) - x, \quad x > 0$$

and

$$h(x) = \log \Gamma(x) - s(x), \quad x > 0. \tag{4.10}$$

Then we have

$$s'(x) = \log x - \frac{1}{2x}, \ s''(x) = \frac{1}{x} + \frac{1}{2x^2}, \ s'''(x) = -\frac{1}{x^2} - \frac{1}{x^3}, \tag{4.11}$$

and there exists a constant $c > 0$ such that

$$|h''(x)| \le \frac{C}{x^3}, \quad x \ge 1/4.$$

Proof. (4.11) is obvious. The last estimate in the lemma follows from (4.9) by noting that $h''(x) = \psi'(x) - s''(x)$. ∎

Lemma 4.5. *We have for any $\delta \ge 1$*

$$0 \le \frac{1}{x^\delta} - \frac{1}{y^\delta} \le \frac{\delta(y - x)}{x^\delta y}, \quad y > x > 0. \tag{4.12}$$

Proof. In fact, for any $\delta \geq 1$, $y > x > 0$

$$0 \leq \frac{1}{x^{\delta}} - \frac{1}{y^{\delta}} = \delta \int_x^y \frac{1}{t^{1+\delta}} dt \leq \frac{\delta}{x^{\delta-1}} \int_x^y \frac{1}{t^2} dt = \frac{\delta}{x^{\delta-1}} (\frac{1}{x} - \frac{1}{y}) = \frac{\delta(y-x)}{x^{\delta} y},$$

proving (4.12). ■

Lemma 4.6. *As $m \to \infty$,*

$$\sum_{i=1}^{q} (\frac{1}{m-i} - \frac{1}{m-1}) = -\frac{q}{m} - \log(1 - \frac{q}{m}) + O(\frac{q}{m(m-q)}), \tag{4.13}$$

$$\sum_{i=1}^{q} \Big(\log(m-i) - \log(m-1) \Big) = (q - m + \frac{1}{2}) \log(1 - \frac{q}{m}) - \frac{(m-1)q}{m} + O(\frac{q}{m(m-q)}), \tag{4.14}$$

$$\sum_{i=1}^{q} (\frac{1}{m-i} - \frac{1}{m-1}) = -\frac{q}{m} - \log(1 - \frac{q}{m}) - \frac{q}{2(m-1)^2} + o(\xi(\frac{q}{m})), \tag{4.15}$$

$$\sum_{i=1}^{q} \frac{1}{(m-i)^2} = \frac{q}{(m-1)^2} + o(\xi(\frac{q}{m})) \tag{4.16}$$

and for any fixed $\delta \geq 2$

$$\sum_{i=1}^{q} \frac{1}{(m-i)^{\delta}} = O(\frac{q}{(m-q)^{\delta-1} m}) \tag{4.17}$$

uniformly over $1 \leq q < m$ as $m \to \infty$.

Proof. (4.13) and (4.14) are proved in Lemma 5 in Qi, Wang and Zhang [18].

Recall the relation of $\xi(x)$ and $\eta(x)$ from (4.2) that $\xi(x) = x^2 \eta(x)$ for $0 < x < 1$. Then we have

$$\frac{q}{m(m-q)} \frac{1}{\xi(\frac{q}{m})} = \frac{m}{q(m-q)} \frac{1}{\eta(\frac{q}{m})}, \quad 1 \leq q < m.$$

Let r_m be any sequence of integers such that $1 \leq r_m < m$ and $r_m/m \to 0$ as $m \to \infty$, From from Lemma 4.1 we have

$$\max_{r_m \leq q < m} \frac{q}{m(m-q)} \frac{1}{\xi(\frac{q}{m})} \to 0 \quad \text{as } m \to \infty$$

and

$$\max_{r_m \leq q < m} \frac{q}{(m-1)^2} \frac{1}{\xi(\frac{q}{m})} \leq \max_{r_m \leq q < m} \frac{q}{m(m-q)} \frac{1}{\xi(\frac{q}{m})} \to 0 \quad \text{as } m \to \infty,$$

which combined with (4.13) yields that (4.15) holds uniformly over $r_m \leq q < m$ as $m \to \infty$. It remains to show that (4.15) holds uniformly over $1 \leq q < r_m$ as $m \to \infty$. In fact, by uisng Taylor's expansion we have uniformly over $1 \le q < r_m$

$$\begin{aligned}
\sum_{i=1}^{q}\left(\frac{1}{m-i}-\frac{1}{m-1}\right) &= \sum_{i=1}^{q}\frac{1}{m-1}\left(\frac{1}{1-\frac{i-1}{m-1}}-1\right)\\
&= \sum_{i=1}^{q}\frac{1}{m-1}\left(\frac{i-1}{m-1}+O\left(\left(\frac{i-1}{m-1}\right)^2\right)\right)\\
&= \sum_{i=1}^{q}\left(\frac{i-1}{(m-1)^2}+O\left(\left(\frac{(i-1)^2}{(m-1)^2}\right)\right)\right)\\
&= \frac{q(q-1)}{2(m-1)^2}+O\left(\frac{q^3}{m^3}\right)\\
&= \frac{q^2}{2(m-1)^2}-\frac{q}{2(m-1)^2}+O\left(\frac{q^3}{m^3}\right)
\end{aligned}$$

and

$$-\frac{q}{m}-\log\left(1-\frac{q}{m}\right)=\frac{q^2}{2m^2}+O\left(\frac{q^3}{m^3}\right),$$

which imply

$$\begin{aligned}
\sum_{i=1}^{q}\left(\frac{1}{m-i}-\frac{1}{m-1}\right) &= -\frac{q}{m}-\log\left(1-\frac{q}{m}\right)-\frac{q}{2(m-1)^2}+\left(\frac{q^2}{(m-1)^2}-\frac{q^2}{m^2}\right)+O\left(\frac{q^3}{m^3}\right)\\
&= -\frac{q}{m}-\log\left(1-\frac{q}{m}\right)-\frac{q}{2(m-1)^2}+O\left(\frac{q^3}{m^3}\right)\\
&= -\frac{q}{m}-\log\left(1-\frac{q}{m}\right)-\frac{q}{2(m-1)^2}+O\left(\frac{r_m}{m}\frac{q^2}{m^2}\right)\\
&= -\frac{q}{m}-\log\left(1-\frac{q}{m}\right)-\frac{q}{2(m-1)^2r}+o\left(\xi\left(\frac{q}{m}\right)\right)
\end{aligned}$$

uniformly over $1 \le q < r_m$ as $m \to \infty$. In the las step we use that $r_m/m \to 0$ as $m \to \infty$ and $\xi(\frac{q}{m}) = \frac{q^2}{m^2}\eta(\frac{q}{m}) \ge \frac{q^2}{m^2}$. This completes the proof of (4.15).

For $1 \le i \le q < m$, we have

$$\begin{aligned}
0 \le \frac{1}{(m-i)^2}-\frac{1}{(m-1)^2} &= \left(\frac{1}{m-i}-\frac{1}{m-1}\right)\left(\frac{1}{m-i}+\frac{1}{m-1}\right)\\
&\le \frac{2}{m-i}\left(\frac{1}{m-i}-\frac{1}{m-1}\right)\\
&= \frac{2(i-1)}{(m-i)^2(m-1)}\\
&\le \frac{2(q-1)}{m-1}\frac{1}{(m-i)^2}\\
&\le \frac{2(q-1)}{m-1}\left(\frac{1}{m-i-0.5}-\frac{1}{m-i+0.5}\right).
\end{aligned}$$

Therefore, from (4.6)

$$\begin{aligned}0 \le \sum_{i=1}^{q} \frac{1}{(m-i)^2} - \frac{q}{(m-1)^2} &\le \frac{2(q-1)}{m-1}\left(\frac{1}{m-q-0.5} - \frac{1}{m-0.5}\right) \\ &= \frac{2(q-1)q}{(m-1)(m-q-0.5)(m-0.5)} \\ &= O\left(\frac{q^2}{m^2(m-q)}\right) \\ &= o\left(\xi\left(\frac{q}{m}\right)\right)\end{aligned}$$

uniformly over $1 \le 1 < m$ as $m \to \infty$, proving (4.16).

When $\delta = 2$, we have from the above proof that

$$\sum_{i=1}^{q} \frac{1}{(m-i)^2} \le \frac{q}{(m-1)^2} + O\left(\frac{q^2}{m^2(m-q)^2}\right) = O\left(\frac{q}{(m-q)m}\right),$$

that is, (4.17) holds with $\delta = 2$. When $\delta > 2$,

$$\sum_{i=1}^{q} \frac{1}{(m-i)^\delta} \le \frac{1}{(m-q)^{\delta-2}} \sum_{i=1}^{q} \frac{1}{(m-i)^2} = O\left(\frac{1}{(m-q)^{\delta-2}} \frac{q}{(m-q)m}\right) = O\left(\frac{q}{(m-q)^{\delta-1}m}\right),$$

proving (4.17). ■

Lemma 4.7. *Assume $p_n \ge 1$ and $k_n \ge 2$ are two sequences of integers such that $1 \le p_n < \min_{1 \le i \le k_n} n_i$ for all large n, and $\min_{1 \le i \le k_n} n_i \to \infty$ as $n \to \infty$. Then*

$$\liminf_{n \to \infty} \frac{\sigma_n^2}{p_n(p_n+1)(k_n-1)} \ge 1, \tag{4.18}$$

and for all large n

$$\eta\left(\frac{p_n}{n-k_n+1}\right) \le \eta\left(\frac{1}{k_n}\right) \le \eta\left(\frac{1}{2}\right), \tag{4.19}$$

where σ_n^2 and $\eta(x)$ are defined in (2.4) and (4.2), respectively.

Proof. Without loss of generality, assume $n_1 = \max_{1 \le i \le k_n} n_i$. Since $\min_{1 \le i \le k_n} n_i > p_n \ge 1$ and $k_n \ge 2$, we have

$$n - k_n + 1 = \sum_{i=1}^{k_n} (n_i - 1) + 1 \ge k_n(n_1 - 1) + 1 > n_1$$

Also note that $\delta_n := 2\max_{1 \le i \le k_n}(1 - \frac{(n_i-1)^2}{n_i^2}) = 2(1 - (1 - \frac{1}{\min_{1 \le i \le k_n} n_i})^2) \to 0$ as $n \to \infty$. From

(4.3), $\eta(\frac{p_n}{n_i}) \geq \eta(\frac{p_n}{n_1}) \geq 1$ and $\eta(\frac{p_n}{n-k_n+1}) \leq \eta(\frac{p_n}{n_1})$. We have for all large n

$$\begin{aligned}
\sigma_n^2 &= p_n^2\Big(\sum_{i=1}^{k_n}\frac{(n_i-1)^2}{n_i^2}\eta(\frac{p_n}{n_i}) - \frac{(n-k_n)^2}{(n-k_n+1)^2}\eta(\frac{p_n}{n-k_n+1})\Big) + p_n(k_n-1) \\
&\geq p_n^2\Big(\sum_{i=1}^{k_n}\frac{(n_i-1)^2}{n_i^2}\eta(\frac{p_n}{n_1}) - \eta(\frac{p_n}{n_1})\Big) + p_n(k_n-1) \\
&\geq p_n^2\Big(\sum_{i=1}^{k_n}\frac{(n_i-1)^2}{n_i^2} - 1\Big)\eta(\frac{p_n}{n_1}) + p_n(k_n-1) \\
&\geq p_n^2\Big(\sum_{i=1}^{k_n}(1-\frac{\delta_n}{2}) - 1\Big) + p_n(k_n-1) \\
&\geq p_n^2(k_n-1)(1-\delta_n) + p_n(k_n-1) \\
&\geq p_n(p_n+1)(k_n-1)(1-\delta_n),
\end{aligned}$$

which implies (4.18).

Since $n - k_n = \sum_{1\leq i\leq k_n}(n_i-1) \geq k_n p_n$, we have

$$\frac{p_n}{n-k_n+1} < \frac{1}{k_n} \leq \frac{1}{2}, \tag{4.20}$$

which yields (4.19) from the monotonicity of η given in (4.3). ■

Lemma 4.8. *Assume there exists a constant $\delta \in (0,1)$ such that*

$$p_n \leq \delta \min_{1\leq i\leq k_n} n_i \text{ for all large } n,\ \lim_{n\to\infty}\frac{\sqrt{k_n}}{\min_{1\leq i\leq k_n} n_i} = 0, \quad \text{and} \lim_{n\to\infty}\min_{1\leq i\leq k_n} n_i = \infty. \tag{4.21}$$

Then

$$\lim_{n\to\infty}\frac{\mu_n - \bar{\mu}_n}{\sigma_n} = 0, \tag{4.22}$$

where μ_n and $\bar{\mu}_n$ are defined in (2.3) *and* (2.10)*, respectively.*

Proof. Define

$$g(x) = \psi(x) - \log x + \frac{1}{2x}, \quad x > 0.$$

Then we have for any positive integers m, q with $m > q \geq 1$

$$\tau(m,q) := \sum_{i=1}^{q} g(\frac{m-i}{2}) = \sum_{i=1}^{q}\psi(\frac{m-i}{2}) - \sum_{i=1}^{q}\log(\frac{m-i}{2}) + \sum_{i=1}^{q}\frac{1}{m-i}. \tag{4.23}$$

Since $n - k_n = \sum_{i=1}^{k_n}(n_i-1) \geq k_n(\min_{1\leq i\leq k_n} n_i - 1) \to \infty$, we have from (4.20) that $n - k_n + 1 - p_n \geq \frac{1}{2}(n-k_n)$ for all large n and $n - k_n - p_n \to \infty$ as $n \to \infty$.

From (4.23), (4.8), condition (4.21) and Lemma 4.7 we have

$$\begin{aligned}
(n-k_n)\tau(n-k_n+1,p_n) &= (n-k_n)\sum_{j=1}^{p_n} g(\frac{n-k_n+1-j}{2}) \\
&= O((n-k_n)\sum_{j=1}^{p_n}\frac{1}{(n-k_n+1-j)^2}) \\
&= O(\frac{p_n}{n-k_n}) \\
&= o(\sigma_n),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{k_n}(n_i-1)\tau(n_i,p_n) &= \sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n} g(\frac{n_i-j}{2}) \\
&= \sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n} O(\frac{1}{(n_i-j)^2}) \\
&= O(p_n)\sum_{i=1}^{k_n}\frac{1}{n_i} \\
&= O(\frac{p_n k_n}{\min_{1\le i\le k_n} n_i}) \\
&= o(p_n\sqrt{k_n}) \\
&= o(\sigma_n).
\end{aligned}$$

Hence we have

$$(n-k_n)\tau(n-k_n+1,p_n)-\sum_{i=1}^{k_n}(n_i-1)\tau(n_i,p_n)=o(\sigma_n).$$

In view of (2.3), (4.23), and equations (4.14) and (4.13) in Lemma 4.6 we have

$$
\begin{aligned}
\mu_n &= (n-k_n)\sum_{j=1}^{p_n}\log(\frac{n-k_n+1-j}{2})-\sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{n_i}\log(\frac{n_i-j}{2}) \\
&\quad -(n-k_n)\sum_{j=1}^{p_n}\frac{1}{n-k_n+1-j}+\sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n}\frac{1}{n_i-j} \\
&\quad +(n-k_n)\tau(n-k_n+1,p_n)-\sum_{i=1}^{k_n}(n_i-1)\tau(n_i,p_n) \\
&\quad +p_n\sum_{i=1}^{k_n}(n_i-1)\log(n_i-1)-p_n(n-k_n)\log(n-k_n) \\
&= (n-k_n)\sum_{j=1}^{p_n}\big(\log(n-k_n+1-j)-\log(n-k_n)\big) \\
&\quad -\sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n}\big(\log(n_i-j)-\log(n_i-1)\big) \\
&\quad -(n-k_n)\sum_{j=1}^{p_n}(\frac{1}{n-k_n+1-j}-\frac{1}{n-k_n})+\sum_{i=1}^{k_n}(n_i-1)\sum_{j=1}^{p_n}(\frac{1}{n_i-j}-\frac{1}{n_i-1}) \\
&\quad +p_n(k_n-1)o(\sigma_n) \\
&= (n-k_n)(p_n-n+k_n-\frac{1}{2})\log(1-\frac{p_n}{n-k_n+1})-\frac{(n-k_n)^2p_n}{n-k_n+1} \\
&\quad -\sum_{i=1}^{k_n}(n_i-1)(p_n-n_i+\frac{1}{2})\log(1-\frac{p_n}{n_i})+\sum_{i=1}^{k_n}\frac{(n_i-1)^2p_n}{n_i} \\
&\quad -(n-k_n)(-\frac{p_n}{n-k_n+1}-\log(1-\frac{p_n}{n-k_n+1}))+\sum_{j=1}^{k_n}(n_i-1)(-\frac{p_n}{n_i}-\log(1-\frac{p_n}{n_i})) \\
&\quad +p_n(k_n-1)+O(\frac{p_n}{n})+O(\sum_{i=1}^{k_n}\frac{p_n}{n_i})+o(\sigma_n) \\
&= (n-k_n)(p_n-n+k_n+\frac{1}{2})\log(1-\frac{p_n}{n-k_n+1})-\sum_{i=1}^{k_n}(n_i-1)(p_n-n_i+\frac{3}{2})\log(1-\frac{p_n}{n_i}) \\
&\quad +\sum_{i=1}^{k_n}\frac{(n_i-1)^2p_n}{n_i}-\frac{(n-k_n)^2p_n}{n-k_n+1}+\frac{(n-k_n)p_n}{n-k_n+1}-\sum_{i=1}^{k_n}\frac{(n_i-1)p_n}{n_i}+p_n(k_n-1)+o(\sigma_n).
\end{aligned}
$$

Moreover, under condition (4.21), we have

$$\begin{aligned}
&\sum_{i=1}^{k_n}\frac{(n_i-1)^2p_n}{n_i}-\frac{(n-k_n)^2p_n}{n-k_n+1}\\
=\;& p_n\sum_{i=1}^{k_n}\frac{(n_i-1)^2-n_i^2}{n_i}-p_n\frac{(n-k_n)^2-(n-k_n+1)^2}{n-k_n+1}+\sum_{i=1}^{k_n}n_ip_n-(n-k_n+1)p_n\\
=\;& p_n\sum_{i=1}^{k_n}(\frac{1}{n_i}-2)-p_n(\frac{1}{n-k_n+1}-2)+p_n(k_n-1)\\
=\;& p_n\sum_{i=1}^{k_n}\frac{1}{n_i}-\frac{p_n}{n-k_n+1}-p_n(k_n-1)\\
=\;& -p_n(k_n-1)+o(\sigma_n)
\end{aligned}$$

and

$$\frac{(n-k_n)p_n}{n-k_n+1}-\sum_{i=1}^{k_n}\frac{(n_i-1)p_n}{n_i}=-p_n(k_n-1)+\sum_{i=1}^{k_n}\frac{p_n}{n_i}-\frac{p_n}{n-k_n+1}=-p_n(k_n-1)+o(\sigma_n).$$

Then we conclude (4.22). ■

Let f be any function defined over $(0,\infty)$. For integers q and m with $1\le q<m$, define the function

$$\Lambda_{f,\,m,q}(t)=\sum_{i=1}^{q}f(\frac{m-i}{2}+t),\quad t>-\frac{m-q}{2}. \tag{4.24}$$

We need this definition in the following proofs.

Lemma 4.9. *For any integers $1\le q<m$, define*

$$G_{m,q}(x)=\sum_{i=1}^{q}\Big(s(\frac{(m-i)}{2}+x)-s(\frac{m-1}{2}+x)\Big),\quad x>\frac{m-q}{2},$$

where the function $s(x)$ is defined in Lemma 4.4. Then

$$\begin{aligned}
&G_{m,q}(\frac{(m-1)t}{2})-G_{m,q}(0) &(4.25)\\
=\;&\Big(\Lambda_{s',\,m,q}(0)-qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2}+\Big(\xi(\frac{q}{m})(1+o(1))-\frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8}
\end{aligned}$$

and

$$\begin{aligned}
&\Lambda_{s,\,m,q}(\frac{(m-1)t}{2})-\Lambda_{s,\,m,q}(0) &(4.26)\\
=\;&\Big(\Lambda_{s',\,m,q}(0)-qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2}+\Big(\xi(\frac{q}{m})(1+o(1))-\frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8}\\
&+\frac{(m-1)q}{2}(1+t)\log(1+t)+\frac{(m-1)q}{2}(1+\log\frac{m-1}{2})t-\frac{q}{2}\log(1+t)
\end{aligned}$$

hold uniformly over $|t|\le(\varepsilon+\frac{1}{m-q})\frac{m-q}{2(m-1)}$, $1\le q<m$ as $\varepsilon\to 0$ and $m\to\infty$.

Proof. We will apply Taylor's theorem to expand a smooth function $f(x)$, that is, that is, for any $x > 0$, $x + y > 0$ there exists a constant c between x and $x + y$ such that

$$f(x+y) = f(x) + f'(x)y + f''(x)\frac{y^2}{2} + f'''(c)\frac{y^3}{6}. \tag{4.27}$$

In our application here, we will assume $x \geq \frac{1}{2}$, and $|y| \leq \frac{1}{2}x$.

We note that

$$\frac{m-i}{2} \geq \frac{1}{4}, \quad \frac{1}{2} \leq \frac{\frac{(m-i)}{2} + \frac{(m-1)t}{2}}{\frac{(m-i)}{2}} \leq \frac{3}{2}, \tag{4.28}$$

when $|t| \leq \frac{m-q}{2(m-1)}$, $1 \leq i \leq q < m$.

Now it follows from (4.27) that for each $1 \leq i \leq q$, there a constant c_i between $\frac{m-i}{2}$ and $\frac{(m-i)}{2} + \frac{(m-1)t}{2}$ such that

$$\begin{aligned} g_i(t) : &= s(\frac{(m-i)}{2} + \frac{(m-1)t}{2}) - s(\frac{m-1}{2} + \frac{(m-1)t}{2}) - (s(\frac{m-i}{2}) - s(\frac{m-1}{2})) \\ &= \big(s'(\frac{m-i}{2}) - s'(\frac{m-1}{2})\big)\frac{(m-1)t}{2} + \big(s''(\frac{m-i}{2}) - s''(\frac{m-1}{2})\big)\frac{(m-1)^2t^2}{8} \\ &\quad + \big(s'''(\frac{(m-i)}{2} + c_i) - s'''(\frac{m-1}{2} + c_i)\big)\frac{(m-1)^3t^3}{48}. \end{aligned}$$

Set

$$R_i = \big(s'''(\frac{(m-i)}{2} + c_i) - s'''(\frac{m-1}{2} + c_i)\big)\frac{(m-1)^3t^3}{48}.$$

Then

$$\begin{aligned} G_{m,q}(\frac{(m-1)t}{2}) &= \sum_{i=1}^{q}(s(\frac{m-i}{2}) - s(\frac{m-1}{2})) + \sum_{i=1}^{q} g_i(t) \\ &= G_{m,q}(0) + \sum_{i=1}^{q} g_i(t) \\ &= G_{m,q}(0) + \Big(\Lambda_{s',\, m,q}(0) - qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2} \\ &\quad + \Big(\Lambda_{s'',\, m,q}(0) - qs''(\frac{m-1}{2})\Big)\frac{(m-1)^2t^2}{8} + \sum_{i=1}^{q} R_i. \end{aligned}$$

From (4.11), (4.28), and Lemma 4.5 we have

$$\begin{aligned} |R_i| &\leq \left(\frac{1}{(\frac{m-i}{2} + c_i)^2} - \frac{1}{(\frac{m-1}{2} + c_i)^2} + \frac{1}{(\frac{m-i}{2} + c_i)^3} - \frac{1}{(\frac{m-1}{2} + c_i)^3}\right)\frac{(m-1)^3|t|^3}{48} \\ &\leq \left(\frac{2i}{(\frac{m-i}{2} + c_i)^2(\frac{m-1}{2} + c_i))} + \frac{3i}{(\frac{m-i}{2} + c_i)^3(\frac{m-1}{2} + c_i))}\right)\frac{(m-1)^3|t|^3}{48} \\ &\leq \frac{64q}{(m-i)^2(m-1)}\frac{(m-1)^3|t|^3}{48} \\ &\leq \frac{2(m-1)^2q|t|^3}{(m-i)^2} \end{aligned}$$

and, thus from (4.17), (4.6), (4.2) and (4.3)

$$
\begin{aligned}
\sum_{i=1}^{q}|R_i| &\le 2(m-1)^2 q|t|^3\sum_{i=1}^{q}\frac{1}{(m-i)^2}\\
&= O(\frac{(m-1)q^2|t|^3}{(m-q)}\\
&= O((\varepsilon+\frac{1}{m-q})\frac{m-q}{2(m-1)}\frac{(m-1)q^2|t|^2}{(m-q)})\\
&= O(\frac{\varepsilon q^2}{(m-1)^2}+\frac{q^2}{(m-1)^2(m-q)})(m-1)^2t^2\\
&= o(\xi(\frac{q}{m}))(m-1)^2t^2
\end{aligned}
$$

holds uniformly over $|t|\le(\varepsilon+\frac{1}{m-q})\frac{m-q}{2(m-1)}$, $1\le q<m$ as $\varepsilon\to 0$ and $m\to\infty$. Furthermore, in view of (4.11), (4.15) and (4.16) we have

$$
\begin{aligned}
\Lambda_{s'',\,m,q}(0)-qs''(\frac{m-1}{2}) &= 2\sum_{i=1}^{q}\Big(\frac{1}{m-i}-\frac{1}{m-1}\Big)+2\sum_{i=1}^{q}\Big(\frac{1}{(m-i)^2}-\frac{1}{(m-1)^2}\Big)\\
&= \xi(\frac{q}{m})-\frac{q}{(m-1)^2}+o(\xi(\frac{q}{m}))
\end{aligned}
$$

uniformly over $1\le q<m$ as $m\to\infty$. (4.25) can be easily obtained by combining all results above.

Review the definition of $s(x)$ in Lemma 4.4 and the definition of Λ in (4.24). Since $\Lambda_{s,\,m,q}(x)=G_{m,q}(x)+qs(\frac{m-1}{2}+x)$, we have

$$
\Lambda_{s,\,m,q}(\frac{(m-1)t}{2})-\Lambda_{s,m,q}(0)=G_{m,q}(\frac{(m-1)t}{2})-G_{m,q}(0)+q\big(s(\frac{(m-1)(1+t)}{2})-s(\frac{m-1}{2})\big).
$$

By using (4.25) and the fact that

$$
\begin{aligned}
& s(\frac{(m-1)(1+t)}{2})-s(\frac{m-1}{2})\\
= & (\frac{(m-1)(1+t)}{2}-\frac{1}{2})\log(\frac{(m-1)(1+t)}{2})-(\frac{m-1}{2}-\frac{1}{2})\log(\frac{m-1}{2})+\frac{(m-1)t}{2}\\
= & \frac{m-1}{2}(1+t)\log(1+t)+\frac{m-1}{2}(1+\log\frac{m-1}{2})t-\frac{1}{2}\log(1+t),
\end{aligned}
$$

we conclude (4.26). ■

Lemma 4.10. *We have*

$$
\begin{aligned}
& \log\frac{\Gamma_q(\frac{m-1}{2}(1+t))}{\Gamma_q(\frac{m-1}{2})}\\
= & \Big(\Lambda_{\psi,\,m,q}(0)-qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2}+\Big(\xi(\frac{q}{m})(1+o(1))-\frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8}\\
& +\frac{(m-1)q}{2}(1+t)\log(1+t)+\frac{(m-1)q}{2}(1+\log\frac{m-1}{2})t-\frac{q}{2}\log(1+t)
\end{aligned}
$$

uniformly over $|t|\le(\varepsilon+\frac{1}{m-q})\frac{m-q}{2(m-1)}$, $1\le q<m$ *as* $\varepsilon\to 0$ *and* $m\to\infty$.

Proof. By the definition of Γ_q function in (4.1), we have

$$\log \frac{\Gamma_q(\frac{m-1}{2}(1+t))}{\Gamma_p(\frac{m-1}{2})} = \sum_{i=1}^{q} \log \frac{\Gamma(\frac{m-i}{2} + \frac{(m-1)t}{2})}{\Gamma(\frac{m-i}{2})} = \Lambda_{\log\Gamma,\, m,q}(\frac{(m-1)t}{2}) - \Lambda_{\log\Gamma,\, m,q}(0). \tag{4.29}$$

The function is well defined when $t > -\frac{m-q}{m-1}$.

For function $h(x)$ defined in (4.10), we apply Taylor's formula with remainder

$$h(\frac{m-i}{2} + \frac{(m-1)t}{2}) - h(\frac{m-i}{2}) = h'(\frac{m-i}{2})\frac{(m-1)t}{2} + h''(c_i)\frac{(m-1)^2t^2}{8} \tag{4.30}$$

where c_i is a number between $\frac{m-i}{2} + \frac{(m-1)t}{2}$ and $\frac{m-i}{2}$. Using Lemma 4.4 we have for some $C > 0$

$$|h''(c_i)| \le \frac{C}{(m-i)^3}, \quad 1 \le i \le q < m < \infty.$$

Summing up over $1 \le i \le q$ on both sides of (4.30) yields

$$\begin{aligned}
\Lambda_{h,\, m,q}(\frac{(m-1)t}{2}) - \Lambda_{h,\, m,q}(0) &= \Lambda_{h',\, m,q}(0)\frac{(m-1)t}{2} + O(\sum_{i=1}^{q}\frac{1}{(m-i)^3})(m-1)^2t^2 \\
&= \Lambda_{h',\, m,q}(0)\frac{(m-1)t}{2} + O(\frac{q}{(m-q)^2m})(m-1)^2t^2 \\
&= \Lambda_{h',\, m,q}(0)\frac{(m-1)t}{2} + o(\xi(\frac{q}{m}))(m-1)^2t^2 \qquad (4.31)
\end{aligned}$$

uniformly over $|t| \le (\varepsilon + \frac{1}{m-q})\frac{m-q}{2(m-1)}$, $1 \le q < m$ as $\varepsilon \to 0$ and $m \to \infty$. In the above estimation, we have used (4.17) from Lemma 4.6 and (4.7) from Lemma 4.2.

From (4.10), we have

$$\Lambda_{\log\Gamma,\, m,q}(\frac{(m-1)t}{2}) = \Lambda_{h,\, m,q}(\frac{(m-1)t}{2}) + \Lambda_{s,\, m,q}(\frac{(m-1)t}{2}).$$

Now by combining (4.29), (4.31), (4.26) and the above equations we have

$$\begin{aligned}
&\log \frac{\Gamma_q(\frac{m-1}{2}(1+t))}{\Gamma_q(\frac{m-1}{2})} \\
=\ & \Lambda_{h,\, m,q}(\frac{(m-1)t}{2}) - \Lambda_{h,\, m,q}(0) + \Lambda_{s,\, m,q}(\frac{(m-1)t}{2}) - \Lambda_{s,\, m,q}(0) \\
=\ & \Lambda_{h',\, m,q}(0)\frac{(m-1)t}{2} + o(\xi(\frac{q}{m}))(m-1)^2t^2 \\
& + \Big(\Lambda_{s',\, m,q}(0) - qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2} + \Big(\xi(\frac{q}{m})(1+o(1)) - \frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8} \\
& \quad + \frac{(m-1)q}{2}(1+t)\log(1+t) + \frac{(m-1)q}{2}(1+\log\frac{m-1}{2})t - \frac{q}{2}\log(1+t) \\
=\ & \Big(\Lambda_{h',\, m,q}(0) + \Lambda_{s',\, m,q}(0) - qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2} + \Big(\xi(\frac{q}{m})(1+o(1)) - \frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8} \\
& \quad + \frac{(m-1)q}{2}(1+t)\log(1+t) + \frac{(m-1)q}{2}(1+\log\frac{m-1}{2})t - \frac{q}{2}\log(1+t) \\
=\ & \Big(\Lambda_{\psi,\, m,q}(0) - qs'(\frac{m-1}{2})\Big)\frac{(m-1)t}{2} + \Big(\xi(\frac{q}{m})(1+o(1)) - \frac{q}{(m-1)^2}\Big)\frac{(m-1)^2t^2}{8} \\
& \quad + \frac{(m-1)q}{2}(1+t)\log(1+t) + \frac{(m-1)q}{2}(1+\log\frac{m-1}{2})t - \frac{q}{2}\log(1+t)
\end{aligned}$$

uniformly over $|t| \le (\varepsilon + \frac{1}{m-q})\frac{m-q}{2(m-1)}$, $1 \le q < m$ as $\varepsilon \to 0$ and $m \to \infty$. This completes the proof of the lemma. ■

Finally, we state the following lemma to express the moments of W_n as a product of Γ_p functions, which is available on page 302 in Muirhead [16].

Lemma 4.11. *Let Λ_n^* be as in (1.2). Set*

$$W_n = \frac{\prod_{i=1}^k |A|^{(n_i-1)/2}}{|A|^{(n-k)/2}} = \Lambda_n^* \cdot \frac{\prod_{i=1}^k (n_i-1)^{(n_i-1)p/2}}{(n-k)^{(n-k)p/2}}. \tag{4.32}$$

Assume $n_i > p$ for $1 \le i \le k$. Then under H_0 in (1.1), we have

$$E(W_n^t) = \frac{\Gamma_p\left(\frac{1}{2}(n-k)\right)}{\Gamma_p\left(\frac{1}{2}(n-k)(1+t)\right)} \cdot \prod_{i=1}^k \frac{\Gamma_p\left(\frac{1}{2}(n_i-1)(1+t)\right)}{\Gamma_p\left(\frac{1}{2}(n_i-1)\right)},$$

for all $t > \max_{1\le i\le k} \frac{p-1}{n_i-1} - 1$, where Γ_p is defined as in (4.1).

4.2 Proofs of the main results

Proof of Theorem 2.1.

We first show the sufficiency, that is, (2.5) holds under condition $\max(p_n, k_n) \to \infty$.

By (4.32), we write

$$-2\log\Lambda_n^* = -2\log W_n + p_n\Big(\sum_{i=1}^{k_n}(n_i-1)\log(n_i-1) - (n-k_n)\log(n-k_n)\Big).$$

Let

$$\mu_n' = \mu_n - p_n\Big(\sum_{i=1}^{k_n}(n_i-1)\log(n_i-1) - (n-k_n)\log(n-k_n)\Big).$$

that

$$\frac{-2\log W_n - \mu_n'}{\sigma_n} \xrightarrow{d} N(0,1),$$

as $n \to \infty$. We will use the moment generating function of $-2\log W_n$ and all we need is to prove that

$$E\exp\left\{\frac{-2\log W_n - \mu_n'}{\sigma_n}w\right\} = E\left(W_n^{-2w/\sigma_n}\right)\exp(\mu_n'(-w/\sigma_n)) \to e^{w^2/2}$$

for all $|w| \le 1/8$. Now we write $t_n = -2w/\sigma_n$. We need to show

$$\lim_{n\to\infty}\left(\log E\left(W_n^{t_n}\right) + \frac{1}{2}\mu_n' t_n\right) = \frac{w^2}{2}.$$

When $|w| \le \frac{1}{8}$, $|t_n| \le \frac{1}{4\sigma_n}$. According to Lemma 4.7, for all large n, $\sigma_n \ge \frac{1}{2}p_n\sqrt{k_n-1}$ and thus $|t_n| \le \frac{1}{2p_n\sqrt{k_n-1}}$. In this case,

$$t_n > -\frac{1}{2p_n\sqrt{k_n-1}} > -\frac{1}{p_n} > \frac{p_n-1}{p_n} - 1 \ge \frac{p_n-1}{\min_{1\le i\le k_n} n_i} - 1 = \max_{1\le i\le k_n}\frac{p_n-1}{n_i} - 1.$$

Therefore, we can apply Lemma 4.11 for all large n to get

$$\log E(W_n^{t_n}) = \sum_{i=1}^{k_n} \log \frac{\Gamma_{p_n}\left(\frac{1}{2}(n_i-1)(1+t_n)\right)}{\Gamma_{p_n}\left(\frac{1}{2}(n_i-1)\right)} - \log \frac{\Gamma_{p_n}\left(\frac{1}{2}(n-k_n)(1+t_n)\right)}{\Gamma_{p_n}\left(\frac{1}{2}(n-k_n)\right)}. \tag{4.33}$$

The following equations will be used in approximating $\log E(W_n^{t_n})$ in (4.33):

$$n - k_n = \sum_{i=1}^{k_n}(n_i - 1),$$

$$\sum_{i=1}^{k_n}(n_i-1)s'(\frac{n_i-1}{2}) - (n-k_n)s'(\frac{n-k_n}{2}) = \sum_{i=1}^{k_n}(n_i-1)\log(n_i-1) - (n-k_n)\log(n-k_n) - (k_n-1), \tag{4.34}$$

$$\frac{(n-k_n)^2p_n^2}{(n-k_n+1)^2}\eta(\frac{p_n}{n-k_n+1}) = O(\sigma_n^2)$$

from Lemma 4.7,

$$\sum_{i=1}^{k_n}(n_i-1)^2\xi(\frac{p_n}{n_i}) \le \sigma_n^2 - \frac{(n-k_n)^2p_n^2}{(n-k_n+1)^2}\eta(\frac{p_n}{n-k_n+1}) = O(\sigma_n^2).$$

and $t_n \to 0$ as $n \to \infty$ since $\max(k_n, p_n) \to \infty$.

For any $m > p_n$,

$$\begin{aligned}
|t_n| &\le \frac{1}{2p_n\sqrt{k_n-1}}\frac{2(m-1)}{m-p_n}\frac{m-p_n}{2(m-1)} \\
&< \frac{1}{\sqrt{k_n-1}}(\frac{m-p_n+p_n)}{(m-p_n}p_n)\frac{m-p_n}{2(m-1)} \\
&= \frac{1}{\sqrt{k_n-1}}(\frac{1}{p_n}+\frac{1}{m-p_n})\frac{m-p_n}{2(m-1)} \\
&= (\frac{1}{p_n\sqrt{k_n-1}}+\frac{1}{m-p_n})\frac{m-p_n}{2(m-1)}.
\end{aligned}$$

Set $\varepsilon_n = \frac{1}{p_n\sqrt{k_n-1}}$. Then $\lim_{n\to\infty}\varepsilon_n = 0$ since $\max(p_n, k_n) \to \infty$. Clearly, we have

$$|t_n| \le \left(\varepsilon_n + \frac{1}{m-p_n}\right)\frac{m-p_n}{2(m-1)}$$

for $m = n_i$, $1 \le i \le k_n$ and $m = n - k_n + 1$. This implies Lemma 4.10 holds uniformly over $m = n_i$,

$1 \le i \le k_n$ and $m = n - k_n$ with $q = p_n$ as $n \to \infty$. Therefore, from (4.33)

$$
\begin{aligned}
& \log E(W_n^{t_n}) \\
= \;& \sum_{i=1}^{k_n} \Big(\Lambda_{\psi,\, n_i, p_n}(0) - p_n s'(\frac{n_i - 1}{2})\Big)\frac{(n_i-1)t_n}{2} - \Big(\Lambda_{\psi,\, n-k_n+1, p_n}(0) - p_n s'(\frac{n-k_n}{2})\Big)\frac{(n-k_n)t_n}{2} \\
& \quad + \sum_{i=1}^{k_n}\Big(\xi(\frac{p_n}{n_i})(1+o(1)) - \frac{p_n}{(n_i-1)^2}\Big)\frac{(n_i-1)^2 t_n^2}{8} \\
& \quad - \Big(\xi(\frac{p_n}{n-k_n+1})(1+o(1)) - \frac{p_n}{(n-k_n)^2}\Big)\frac{(n-k_n)^2 t_n^2}{8} \\
& \quad + \sum_{i=1}^{k_n}\Big(\frac{(n_i-1)p_n}{2}(1+t_n)\log(1+t_n) + \frac{(n_i-1)p_n}{2}(1+\log\frac{n_i-1}{2})t_n - \frac{p_n}{2}\log(1+t_n)\Big) \\
& \quad - \Big(\frac{(n-k_n)p_n}{2}(1+t_n)\log(1+t_n) + \frac{(n-k_n)p_n}{2}(1+\log\frac{n-k_n}{2})t_n - \frac{p_n}{2}\log(1+t_n)\Big) \\
= \;& \frac{t_n}{2}\Big\{\sum_{i=1}^{k_n}(n_i-1)\Lambda_{\psi,\, n_i, p_n}(0) - (n-k_n)\Lambda_{\psi,\, n-k_n+1, p_n}(0)\Big\} \\
& \quad - \frac{p_n t_n}{2}\Big\{\sum_{i=1}^{k_n}(n_i-1)s'(\frac{n_i-1}{2}) - (n-k_n)s'(\frac{n-k_n}{2}))\Big\} \\
& \quad + \frac{t_n^2}{8}\Big\{\sum_{i=1}^{k_n}(n_i-1)^2\xi(\frac{p_n}{n_i}) - (n-k_n)^2\xi(\frac{p_n}{n-k_n+1}) - p_n(k_n-1)\Big\}(1+o(1)) \\
& \quad + \frac{p_n t_n}{2}\Big\{\sum_{i=1}^{k_n}(n_i-1)\log\frac{n_i-1}{2} - (n-k_n)\log\frac{n-k_n}{2}\Big\} \\
& \quad - \frac{p_n(k_n-1)}{2}\log(1+t_n) \\
= \;& : I_1 + I_2 + I_3 + I_4 + I_5.
\end{aligned}
$$

Since

$$I_1 = -\frac{1}{2}\mu_n' t_n,$$

$$I_2 + I_4 = \frac{1}{2}p_n(k_n-1)t_n$$

from (4.34), and

$$I_5 = -\frac{p_n(k_n-1)}{2}\log(1+t_n) = -\frac{1}{2}p_n(k_n-1)t_n + p_n(k_n-1)\frac{t_n^2}{4}(1+o(1)).$$

by Taylor's expansion, we have

$$
\begin{aligned}
& \log E(W_n^{t_n}) + \frac{1}{2}\mu_n' t_n \\
= \quad & \frac{t_n^2}{8}\Big(\sum_{i=1}^{k_n}(n_i-1)^2)\xi(\frac{p_n}{n_i}) - (n-k_n)^2\xi(\frac{p_n}{n-k_n+1}) + p_n(k_n-1)\Big)(1+o(1)) \\
= \quad & \frac{1}{8}\Big(\frac{-2w}{\sigma_n}\Big)^2\sigma_n^2(1+o(1)) \\
\to \quad & \frac{w^2}{2}
\end{aligned}
$$

as $n\to\infty$.

To prove the necessity, we need to show that (2.5) implies $\lim_{n\to\infty}\max(p_n,k_n)=\infty$. If condition $\lim_{n\to\infty}\max(p_n,k_n)=\infty$ is not sure, then we see that exists a subsequence of n, say, $\{n'\}$ such that $p_{n'}=p$ and $k_{n'}=k$ for all large n', where p and k are two fixed integers. Then it follows from the classic chi-square approximation (1.3) that $-2\log\Lambda_{n'}^*$ converges in distribution to a chi-square distribution with $f=\frac{1}{2}p(p+1)(k-1)$ degrees of freedom as $n'\to\infty$. This is contradictory to (2.5).

This completes the proof of the theorem. ■

Proof of Theorem 2.2.

To prove (2.8), it suffices to show that for any subsequence $\{n'\}$ of $\{n\}$, there is a further subsequence $\{n''\}$ such that (2.8) holds along $\{n''\}$. The subsequence $\{n''\}$ can be selected in a way that both the limits of $k_{n''}$ and $p_{n''}$ exist, and both the limits can be infinity. For the sake of simplicity, we can assume both the limits of k_n and p_n exist along the entire sequence and prove (2.8) holds. Note that if the limit of a sequence of integers is finite, the sequence takes a constant value ultimately. For this reason, it suffices to show (2.8) under conditions $\min_{1\le i\le k_n} n_i > p_n$ and $\min_{1\le i\le k_n} n_i\to\infty$, and each of the following two assumptions:

Assumption 1: $\lim_{n\to\infty}\max(p_n,k_n)=\infty$;

Assumption 2: $p_n=p$ and $k_n=k$ for all large n, where both p and k are fixed integers.

Assume Assumption 1 holds. In this case, Theorem 2.1 holds, and $f_n=\frac{1}{2}p_n(p_n+1)(k_n-1)\to\infty$ as $n\to\infty$. By using Theorem 2.1 and applying the central limit theorem to $\chi^2_{f_n}$ we have

$$\sup_{-\infty<x<\infty}|P(\frac{-2\log\Lambda_n-\mu_n}{\sigma_n}\le x)-\Phi(x)|\to 0 \text{ and } \sup_{-\infty<x\infty}|P(\frac{\chi^2_{f_n}-f_n}{\sqrt{2f_n}}\le x)-\Phi(x)|\to 0$$

as $n\to\infty$, where $\Phi(x)$ denotes the cumulative distribution function for the standard normal random variable. Therefore,

$$\begin{aligned}
& \sup_{-\infty<x<\infty} |P(Z_n \le x) - P(\chi^2_{f_n} \le x)| \\
= & \sup_{-\infty<x<\infty} |P(\frac{Z_n - f_n}{\sqrt{2f_n}} \le x) - P(\frac{\chi^2_{f_n} - f_n}{\sqrt{2f_n}} \le x)| \\
\le & \sup_{-\infty<x<\infty} |P(\frac{Z_n - f_n}{\sqrt{2f_n}} \le x) - \Phi(x)| + \sup_{-\infty<x<\infty} |P(\frac{\chi^2_{f_n} - f_n}{\sqrt{2f_n}} \le x) - \Phi(x)| \\
\to & \ 0 \quad \text{as } n \to \infty,
\end{aligned}$$

proving (2.8).

Under Assumption 2, we have that $f_n = f = \frac{1}{2}p(p+1)(k-1)$, as defined in (1.4), is a fixed integer. It suffices to show that Z_n converges in distribution to χ^2_f. From (2.7), if we are able to show

$$\frac{2f_n}{\sigma_n^2} \to 1 \quad \text{and} \quad f_n - \mu_n\sqrt{\frac{2f_n}{\sigma_n^2}} \to 0 \tag{4.35}$$

as $n \to \infty$, then $Z_n = (-2\log\Lambda_n^*)(1+o(1)) + o(1)$, which concludes the desired result by using (1.3).

It is easy to see that

$$\begin{aligned}
\lim_{n\to\infty} \sigma_n^2 &= \lim_{n\to\infty} \Big(\sum_{i=1}^k (n_i-1)^2 (\frac{p}{n_i})^2 \eta(\frac{p}{n_i}) - (n-k_n)^2 (\frac{p}{n-k_n})^2 \eta(\frac{p}{n-k_n+1}) \Big) + p(k-1) \\
&= \sum_{i=1}^k p^2\eta(0) - p^2\eta(0) + p(k-1) \\
&= p^2(k-1) + p(k-1) \\
&= p(p+1)(k-1).
\end{aligned}$$

Next, we need to estimate μ_n. Since Assumption 2 implies condition (4.21), we can estimate $\bar{\mu}_n$ and then apply (4.22). By using Taylor's expansion for $\bar{\mu}_n$ given in (2.10), we can show $\bar{\mu}_n = \frac{1}{2}p(p+1)(k-1)+o(1)$. We omit the details here. Therefore, we have $\mu_n = \bar{\mu}_n + \sigma_n(\frac{\mu_n - \bar{\mu}}{\sigma_n}) \to \frac{1}{2}p(p+1)(k-1)$ as $n \to \infty$. Then (4.35) follows easily from the limits of σ_n^2 and μ_n. ■

Proof of Theorem 2.3. Part of (b) of the theorem follows from Theorem 2.1 and equation (4.22) in Lemma 4.8. Part (a) follows from the same lines in the proof of Theorem 2.2 by using Part (b) of Theorem 2.3. We omit the details. ■

Supplementary Material

The Supplementary Material contains Figures S2 and S3.

Acknowledgements

The authors would like to thank the Editor, Associate Editor, and referees for reviewing the manuscript and providing valuable comments. The research of Yongcheng Qi was supported in part by NSF Grant DMS-1916014.

Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] M. Abramowitz and I.A. Stegun (1972). Handbook of Mathematical Functions. *Dover, New York*.

[2] T. W. Anderson (1958). An introduction to multivariate statistical analysis. Wiley Publications in Statistics. John Wiley & Sons, New York.

[3] Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. Ann. Statist. 37(6B), 3822–3840.

[4] M. S. Bartlett (1937). Properties of sufficiency and statistical tests. Proc. R. Soc. A 160(901), 268–282.

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

[6] G. W. Brown (1939). On the power of the L_1 test for equality of several variances. Ann. Math. Statistics 10(2), 119–128.

[7] T. T. Cai, W. D. Liu and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high dimensional and sparse settings. J. Amer. Statist. Assoc. 108, 265–277.

[8] H. Dette and N. Dörnemann (2020). Likelihood ratio tests for many groups in high dimensions. Journal of Multivarite Analysis 178, 104605.

[9] M. L. Eaton (1983). Multivariate statistics: A Vector Space Approach. Wiley, New York.

[10] W. Guo (2014). Central limit theorem for testing the equality of covariance matrices. Master thesis, University of Minnesota Duluth.
https://scse.d.umn.edu/sites/scse.d.umn.edu/files/guotr.pdf

[11] D. Jiang, T. Jiang, and F. Yang (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. J. Statist. Plann. Inference 142(8), 2241–2256.

[12] T. Jiang and Y. Qi (2015). Likelihood ratio tests for high-dimensional normal distributions. Scandinavian Journal of Statistics 42, 988–1009.

[13] T. Jiang and F. Yang (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. Ann. Statist. 41(4), 2029–2074.

[14] O. Ledoit and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Statist. 30(4), 1081–1102.

[15] J. Li and S. X. Chen (2012). Two sample tests for high dimensional covariance matrices. Ann. Statist 40, 908–940.

[16] R. J. Muirhead (1982). Aspects of multivariate statistical theory. John Wiley & Sons, New York, Wiley Series in Probability and Mathematical Statistics.

[17] M. D. Perlman (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. Ann. Statist. 8(2), 247–263.

[18] Y. Qi, F. Wang and L. Zhang (2019) Likelihood ratio test of independence of components for high-dimensional normal vectors. Annals of the Institute of Statistical Mathematics 71, 911–946.

[19] J. R. Schott (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.

[20] M. S. Srivastava and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. J. Multivariate Anal. 101, 1319–1329.

[21] N. Sugiura and H. Nagao (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. Ann. Math. Statist. 39, 1686-1692.

[22] S. S. Wilks (1932). Certain generalizations in the analysis of variance. Biometrika 24(3/4), 471–494.

[23] Q. Yang and G. Pan (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. J. Amer. Statist. Assoc. 517, 188–200.

[24] S. Zheng, R. Lin, J. Guo and G. Yin (2020). Testing homogeneity of high-dimensional covariance matrices. Statistica Sinica 30, 35–53.

Asymptotic Distributions for Likelihood Ratio Tests for the Equality of Covariance Matrices

Wenchuan Guo[a], **Yongcheng Qi**[b]

[a]Biometrics & Data Sciences, Bristol Myers Squibb, 3551 Lawrenceville Princeton, Lawrence Township, NJ 08543, USA.

[b]Department of Mathematics and Statistics, University of Minnesota Duluth, 1117 University Drive, Duluth, MN 55812, USA.

Supplementary Material

We present additional plots for histograms of $-2\rho\log\Lambda_n^*$ (Chisq), $(-2\log\Lambda_n^* - \mu_n)/\sigma_n$ (CLT), and Z_n (ALRT).

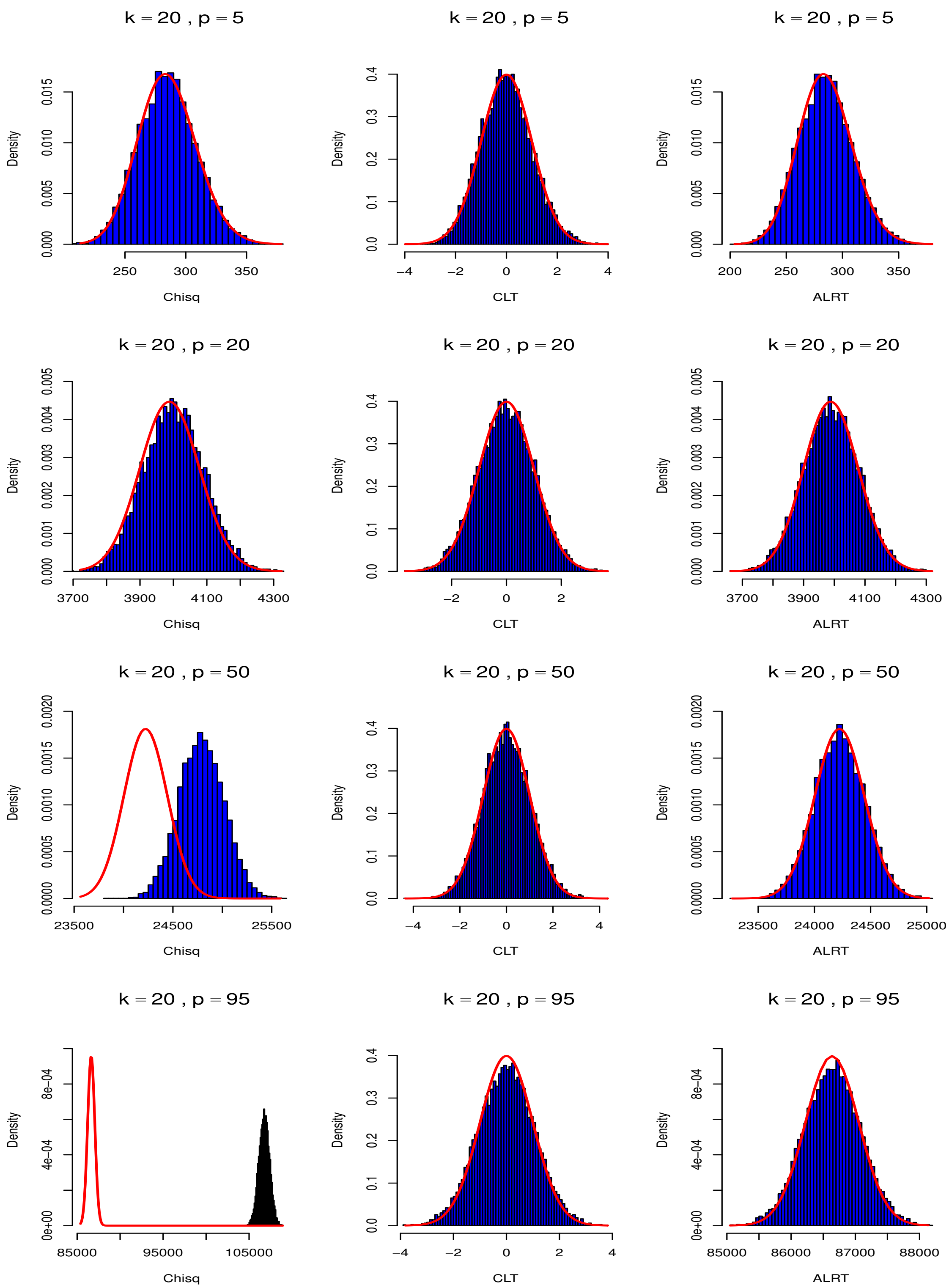

Figure S2: Histograms of $-2\rho \log \Lambda_n^*$ (Chisq), $(-2 \log \Lambda_n^* - \mu_n)/\sigma_n$ (CLT), and Z_n (ALRT), where $n_i = 100$, $1 \leq i \leq k$, with $k = 20$ and $p = 5, 20, 50, 95$.

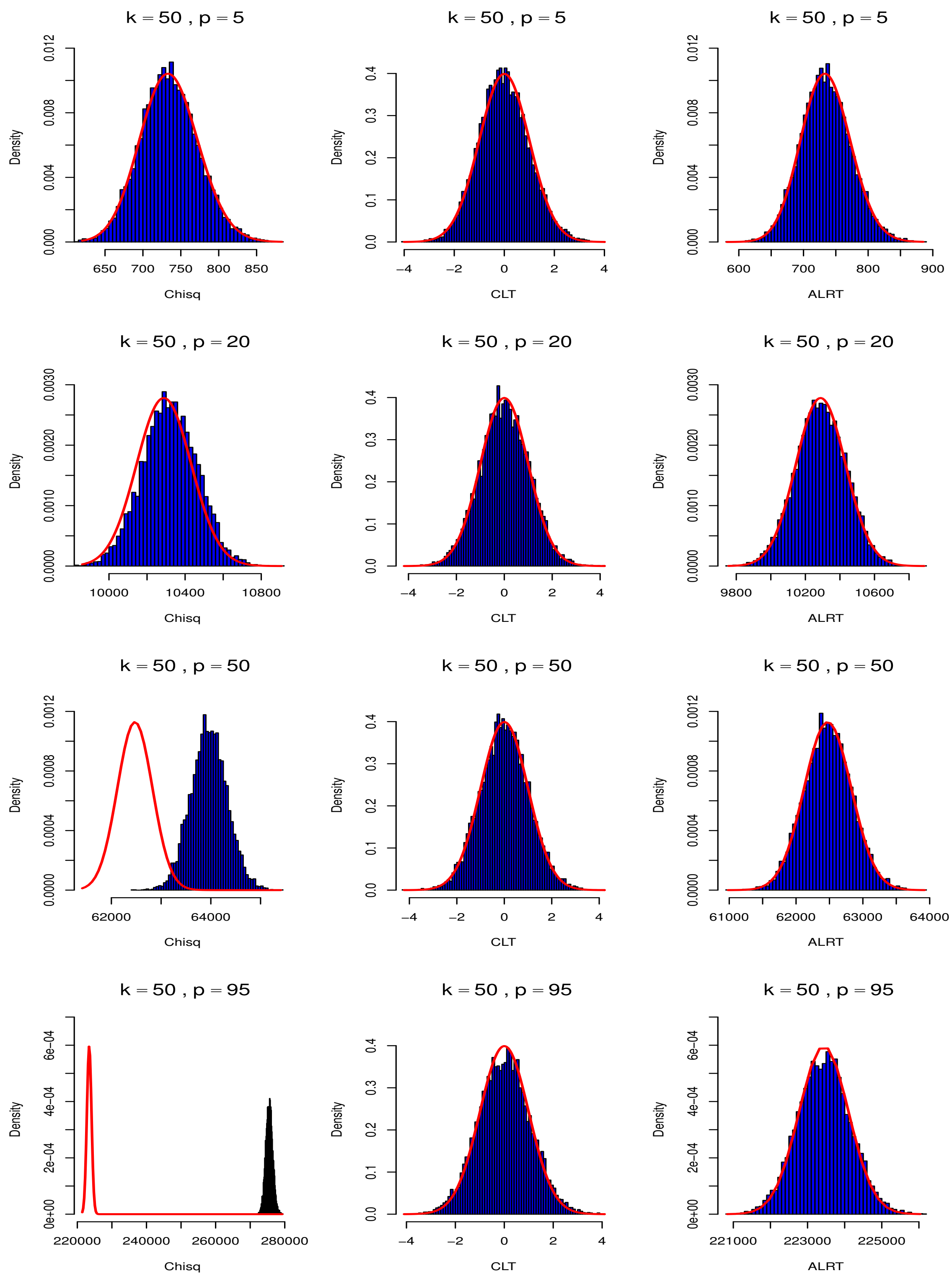

Figure S3: Histograms of $-2\rho \log \Lambda_n^*$ (Chisq), $(-2 \log \Lambda_n^* - \mu_n)/\sigma_n$ (CLT), and Z_n (ALRT), where $n_i = 100$, $1 \le i \le k$, with $k = 50$ and $p = 5, 20, 50, 95$.