Journal of Multivariate Analysis 141 (2015) 217–227

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A robust test for sphericity of high-dimensional covariance matrices

Xintao Tian [a,b], Yuting Lu [c], Weiming Li [a,*]

[a] School of Science, Beijing University of Posts and Telecommunications, China
[b] School of Statistics, Renmin University of China, China
[c] School of Mathematics, Beijing Normal University, China

ARTICLE INFO

Article history:
Received 11 April 2015
Available online 5 August 2015

AMS 2010 subject classifications:
62H15
62H10

Keywords:
Covariance matrix
High-dimensional data
Sphericity test

ABSTRACT

This paper discusses the problem of testing the sphericity of a covariance matrix in high-dimensional frameworks. A new test procedure is put forward by taking the maximum of two existing statistics which are proved weakly independent in our settings. Asymptotic distribution of the new statistic is derived for generally distributed population with a finite fourth moment. Extensive simulations demonstrate that the proposed test has a great improvement in robustness of power against various models under the alternative hypothesis.

1. Introduction

Large-scale statistical inferences involving covariance matrices are increasingly encountered in many scientific research fields, such as signal processing, image processing, genetics, and stock marketing. A basic problem among such inferences is the sphericity test for covariance matrices when the number of observations is not negligible with respect to the sample size.

Generally, let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^p$, be a sequence of independent and identically distributed zero-mean random vectors with a common population covariance matrix Σ_p. The sample covariance matrix takes the form $S_n = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'/n$. Hypotheses regarding to the sphericity test are

$$H_0 : \Sigma_p = \sigma^2 I_p \quad \text{vs.} \quad H_1 : \Sigma_p \neq \sigma^2 I_p,$$

where σ^2 is the unknown scalar proportion. Our interest is to study this test based on S_n for general population, in an asymptotic framework where both p and n tend to infinity with $p/n \to c \in (0, \infty)$.

There are a number of works in the literature addressing the sphericity test in high-dimensions. Ledoit and Wolf [12] generalized the locally best invariant (LBI) test which was set up by [9,10] in a fixed p context. This result was later refined in [18] by applying an unbiased estimator of $\mathrm{tr}(\Sigma_p^2)/p$. Fisher et al. [8] studied a homogeneous test constructed from unbiased estimators of $\mathrm{tr}(\Sigma_p^k)/p$, $k = 2, 4$. However, these tests heavily rely on an assumption that the sample is normally distributed. For non-normal cases, Chen et al. [6] developed a new method where the statistic was constituted by some well selected U-statistics, but this technique carries a burden of doing extensive computations. Srivastava et al. [20] proved that the test in [18] is still valid when the kurtosis of the underlying distribution is close to 3, the Gaussian case. Recently, Wang and

* Corresponding author.
E-mail addresses: xttian@ruc.edu.cn (X. Tian), luyut99@sina.com (Y. Lu), liwm@bupt.edu.cn (W. Li).

http://dx.doi.org/10.1016/j.jmva.2015.07.010
0047-259X/

Yao [21] corrected the LBI test and the Gaussian likelihood ratio test [1] in high-dimensions for arbitrary distribution, and provided a finite fourth moment. For more references, one is referred to [11,19,5,15,4], etc.

Among all these tests, we are particularly interested in those from [18,8]. Asymptotic joint distribution of these two test statistics is derived for general populations under both the null and alternative hypotheses. With this joint distribution, we are surprised to find that the two tests are almost independent when the limiting ratio c is large, so that one test statistic has significant changes while the other may not, and vice versa. This inspires us to propose a new test procedure that can reject the sphericity hypothesis when any of the two statistics is large. It turns out that the new test achieves excellent performance on robustness of power compared with the original ones.

The rest of the paper is organized as follows. Section 2 reviews the unbiased estimators of $\mathrm{tr}(\Sigma_p^k)/p$ built on normality assumption and investigates their asymptotic behaviors for general population under moment conditions. In Section 3, we discuss the relationship between Srivastava's test and Fisher et al.'s test, and then formulate our new test procedure. Section 4 reports simulation results and Section 5 presents conclusions and remarks. Technical proofs are deferred to the last section.

2. Estimators of $\mathrm{tr}(\Sigma_p^k)/p$ and their asymptotic properties

Let H_p and F_n be spectral distributions of Σ_p and S_n, respectively. Then integer moments of H_p and F_n are defined by

$$\alpha_k := \int t^k dH_p(t) = \frac{1}{p}\mathrm{tr}(\Sigma_p^k) \quad \text{and} \quad \hat{\beta}_k := \int x^k dF_n(x) = \frac{1}{p}\mathrm{tr}(S_n^k),$$

$k = 0, 1, 2, \ldots$. Assuming the sample data are normally distributed, estimators of $\alpha_i,\ i = 1, 2, 3, 4$, employed in succession in [18,8,7] were proved to be unbiased, consistent, and asymptotically normal. Moreover, these estimators can be expressed as polynomials of $\hat{\beta}_k$'s, i.e.,

$$\begin{aligned}
\hat{\alpha}_1 &= \hat{\beta}_1,\\
\hat{\alpha}_2 &= \tau_2\left(\hat{\beta}_2 - c_n\hat{\beta}_1^2\right),\\
\hat{\alpha}_3 &= \tau_3\left(\hat{\beta}_3 - 3c_n\hat{\beta}_2\hat{\beta}_1 + 2c_n^2\hat{\beta}_1^3\right),\\
\hat{\alpha}_4 &= \tau_4\left(\hat{\beta}_4 - 4c_n\hat{\beta}_3\hat{\beta}_1 - \frac{2n^2+3n-6}{n^2+n+2}c_n\hat{\beta}_2^2 + \frac{10n^2+12n}{n^2+n+2}c_n^2\hat{\beta}_2\hat{\beta}_1^2 - \frac{5n^2+6n}{n^2+n+2}c_n^3\hat{\beta}_1^4\right),
\end{aligned}$$

where $c_n = p/n,\ \tau_2 = n^2/[(n-1)(n+2)],\ \tau_3 = n^4/[(n-1)(n-2)(n+2)(n+4)]$, and $\tau_4 = n^5(n^2+n+2)/[(n+1)(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)]$. When the underlying distribution is not normal, we show that the unbiasedness does not hold any more for $\hat{\alpha}_2, \hat{\alpha}_3$, and $\hat{\alpha}_4$, but the consistency and asymptotic normality can be retained under suitable assumptions.

Assumption (a). The sample and population sizes n, p both tend to infinity, in such a way that $c_n = p/n \to c \in (0, \infty)$.

Assumption (b). There is a doubly infinite array of i.i.d. random variables $(w_{ij}),\ i,\ j \geq 1$, satisfying

$$E(w_{11}) = 0, \qquad E(w_{11}^2) = 1, \qquad E(w_{11}^4) < \infty,$$

such that for each p, n, letting $W_n = (w_{ij})_{1\le i\le p, 1\le j\le n}$, the observation vectors can be represented as $\mathbf{x}_j = \Sigma_p^{1/2} w_{\cdot j}$ where $w_{\cdot j} = (w_{ij})_{1\le i\le p}$ denotes the jth column of W_n.

Assumption (c). The population spectral distribution H_p of Σ_p weakly converges to a probability distribution H, as $p \to \infty$, and the sequence of spectral norms $(\|\Sigma_p\|)$ is bounded.

Assumptions (a)–(c) are classical conditions of the central limit theorem for linear spectral statistics of sample covariance matrices, see [2,3]. From the third assumption, moments of H_p converge to the corresponding moments of H, that is,

$$\alpha_k \to \tilde{\alpha}_k := \int t^k dH(t),$$

as $p \to \infty$, for any fixed $k \in \mathbb{N}$.

Lemma 1. *Suppose that Assumptions (a)–(c) hold, then*

(i) *the estimator $\hat{\alpha}_k$ is strongly consistent, i.e.,*

$$\hat{\alpha}_k - \alpha_k \xrightarrow{a.s.} 0, \quad k = 1, 2, 3, 4.$$

(ii) *If in addition* $E(w_{11}^4) = 3$, *then*

$$n\left(\hat{\alpha}_1 - \alpha_1, \hat{\alpha}_2 - \alpha_2, \hat{\alpha}_3 - \alpha_3, \hat{\alpha}_4 - \alpha_4\right)' \xrightarrow{D} N_4(0, A), \tag{1}$$

where the covariance matrix is given in Appendix A.1.

Lemma 1 states strong consistency and asymptotic normality of the estimators $\hat{\alpha}_k$'s. The convergence in (1) is the same as Theorem 2 in [7] which assumed w_{11} is standard normal, while our result only requires that w_{11} matches a standard normal variable at the first, second and fourth moments. If the fourth moment is not equal to 3, then the distribution in (1) may have a large change. We present below such a result which is also the cornerstone of our test procedure.

Lemma 2. *In addition to Assumptions (a)–(c), suppose that* Σ_p *is diagonal for all p large. Then*

$$n\left(\hat{\alpha}_1 - \alpha_1, \hat{\alpha}_2 - \alpha_2, \hat{\alpha}_3 - \alpha_3, \hat{\alpha}_4 - \alpha_4\right)' \xrightarrow{D} N_4(v, B),$$

where the mean vector v *and the covariance matrix B are respectively*

$$v = \Delta \cdot \begin{pmatrix} 0 \\ \tilde{\alpha}_2 \\ 3\tilde{\alpha}_3 \\ 6\tilde{\alpha}_4 \end{pmatrix} \quad \text{and} \quad B = A + \frac{\Delta}{c} \cdot \begin{pmatrix} \tilde{\alpha}_2 & 2\tilde{\alpha}_3 & 3\tilde{\alpha}_4 & 4\tilde{\alpha}_5 \\ 2\tilde{\alpha}_3 & 4\tilde{\alpha}_4 & 6\tilde{\alpha}_5 & 8\tilde{\alpha}_6 \\ 3\tilde{\alpha}_4 & 6\tilde{\alpha}_5 & 9\tilde{\alpha}_6 & 12\tilde{\alpha}_7 \\ 4\tilde{\alpha}_5 & 8\tilde{\alpha}_6 & 12\tilde{\alpha}_7 & 16\tilde{\alpha}_8 \end{pmatrix} \tag{2}$$

with the constant $\Delta = E(w_{11}^4) - 3$ *and the matrix A defined in (1).*

Lemma 2 presents a new central limit theorem for the estimators $\hat{\alpha}_k$'s under general fourth moment of w_{11}, assuming Σ_p is diagonal. One can see from (2) that both the mean vector and the covariance matrix create shifts from those in (1), which meanwhile denies the unbiasedness of $\hat{\alpha}_k$ for $k = 2, 3, 4$. When neither $\Delta = 0$ nor Σ_p is diagonal, the limiting distribution will become more complex, depending on the eigenvectors of Σ_p, see [16].

3. Test procedure

From the Cauchy–Schwarz inequality, moments of H_p satisfy

$$d_k := \frac{\alpha_{2k}}{\alpha_k^2} - 1 \geq 0, \quad k \in \mathbb{N}. \tag{3}$$

The equality in (3) holds if and only if the hypothesis of sphericity stands, which is often used in the sphericity test. Srivastava [18] tested if $d_1 = 0$, while Fisher et al. [8] examined if $d_2 = 0$. Their statistics are respectively

$$T_s := \frac{\hat{\alpha}_2}{\hat{\alpha}_1^2} - 1 \quad \text{and} \quad T_f := \frac{\hat{\alpha}_4}{\hat{\alpha}_2^2} - 1,$$

which are consistent estimators of d_1 and d_2.

Theorem 1. *Suppose that Assumptions (a)–(c) hold.*

(i) *If* Σ_p *is diagonal for all p large, then*

$$n\left(T_s - d_1, T_f - d_2\right)' \xrightarrow{D} N_2(\mu, \Omega), \tag{4}$$

as $(n, p) \to \infty$, *where the mean vector is* $\mu = \Delta \cdot (\tilde{\alpha}_2/\tilde{\alpha}_1^2, 4\tilde{\alpha}_4/\tilde{\alpha}_2^2)'$ *with* $\Delta = E(w_{11}^4) - 3$ *and the covariance matrix* $\Omega = (\omega_{ij})$ *with its entries*

$$\omega_{11} = \frac{4}{c\tilde{\alpha}_1^6}\left(2\tilde{\alpha}_2^3 - 4\tilde{\alpha}_1\tilde{\alpha}_2\tilde{\alpha}_3 + 2\tilde{\alpha}_1^2\tilde{\alpha}_4 + c\tilde{\alpha}_1^2\tilde{\alpha}_2^2 + \Delta \cdot (\tilde{\alpha}_2^3 - 2\tilde{\alpha}_1\tilde{\alpha}_2\tilde{\alpha}_3 + \tilde{\alpha}_1^2\tilde{\alpha}_4)\right),$$

$$\begin{aligned}\omega_{12} = {} & \frac{8}{c\tilde{\alpha}_1^3\tilde{\alpha}_2^3}\left(2\tilde{\alpha}_1\tilde{\alpha}_2\tilde{\alpha}_6 + c\tilde{\alpha}_1\tilde{\alpha}_2\tilde{\alpha}_3^2 + c\tilde{\alpha}_1\tilde{\alpha}_2^2\tilde{\alpha}_4 - 2\tilde{\alpha}_1\tilde{\alpha}_4^2 + 2\tilde{\alpha}_2\tilde{\alpha}_3\tilde{\alpha}_4 - 2\tilde{\alpha}_2^2\tilde{\alpha}_5\right. \\ & \left. + \Delta \cdot (\tilde{\alpha}_2\tilde{\alpha}_3\tilde{\alpha}_4 - \tilde{\alpha}_1\tilde{\alpha}_4^2 - \tilde{\alpha}_2^2\tilde{\alpha}_5 + \tilde{\alpha}_1\tilde{\alpha}_2\tilde{\alpha}_6)\right),\end{aligned}$$

$$\begin{aligned}\omega_{22} = {} & \frac{8}{c\tilde{\alpha}_2^6}\left(4\tilde{\alpha}_4^3 - 8\tilde{\alpha}_2\tilde{\alpha}_4\tilde{\alpha}_6 + 4\tilde{\alpha}_2^2\tilde{\alpha}_8 + 8c\tilde{\alpha}_2^2\tilde{\alpha}_3\tilde{\alpha}_5 + 4c\tilde{\alpha}_2^3\tilde{\alpha}_6 - 4c\tilde{\alpha}_2\tilde{\alpha}_3^2\tilde{\alpha}_4\right. \\ & \left. + 8c^2\tilde{\alpha}_2^3\tilde{\alpha}_3^2 + 4c^2\tilde{\alpha}_2^4\tilde{\alpha}_4 + c^3\tilde{\alpha}_2^6 + \Delta \cdot (2\tilde{\alpha}_4^3 - 4\tilde{\alpha}_2\tilde{\alpha}_4\tilde{\alpha}_6 + 2\tilde{\alpha}_2^2\tilde{\alpha}_8)\right).\end{aligned}$$

(ii) *If* $E(w_{11}^4) = 3$ *then the convergence in (4) also holds with the same limiting mean vector (actually zero vector) and covariance matrix.*

(iii) *Under the null hypothesis,*

$$n\left(T_s, T_f\right)' \xrightarrow{D} N_2(\mu_0, \Omega_0),$$

as $(n, p) \to \infty$, *where the mean vector is* $\mu_0 = (\Delta, 4\Delta)'$ *and the covariance matrix* $\Omega_0 = (\omega_{0ij})$ *with* $\omega_{011} = 4,\ \omega_{012} = 16$, *and* $\omega_{022} = 8(8 + 12c + c^2)$.

Theorem 1 illustrates asymptotic joint distribution of T_s and T_f under various cases. Conclusions (i) and (ii) reveal that T_s and T_f become less dependent as the dimensional ratio c gets larger, because their limiting correlation coefficient $\omega_{12}/\sqrt{\omega_{11}\omega_{22}}$ is roughly proportional to $1/c$. This fact demonstrates the two tests can complement and reinforce each other in high-dimensional frameworks, especially when c is large. Conclusion (iii) shows that the original test procedures built on normality assumption are not applicable to general population with $E(w_{11}^4) \neq 3$, as they may suffer from serious size distortion. Note that the asymptotic null distribution of T_s has already been obtained in [21] as the null distribution of their corrected John's test.

Based on these findings, we present our test statistic T_m as the maximum of standardized T_s and T_f, that is,

$$T_m := \max\left\{\frac{nT_s - \Delta}{2}, \frac{nT_f - 4\Delta}{\sqrt{8(8 + 12c_n + c_n^2)}}\right\}.$$

Notice that the two components in T_m are both asymptotically standard normal under the sphericity hypothesis. Taking their maximum demonstrates our test will reject the sphericity when any of them is large.

Theorem 2. *Suppose that Assumptions (a)–(c) hold.*

(i) *Under the null hypothesis, for any* $x \in \mathbb{R}$,

$$P\left(T_m \leq x\right) \to \int_{-\infty}^{x}\int_{-\infty}^{x} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right\} du dv,$$

as $(n, p) \to \infty$, *where* $\rho = 4/\sqrt{2(8 + 12c + c^2)}$.

(ii) *If the spectral distribution* $H \neq \delta_{\sigma^2}$, *a Dirac point measure at* σ^2, *then, for any* $x \in \mathbb{R}$,

$$P\left(T_m > x\right) \to 1, \quad \text{as } (n, p) \to \infty.$$

This theorem is a direct result of Theorem 1, where the first conclusion sets forth the asymptotic null distribution of our test and the second guarantees its consistency.

4. Simulation study

We numerically evaluate finite-sample performance of the corrected Srivastava's test and Fisher et al.'s test, still denoted by T_s and T_f, and our test T_m. The test in [6] and the corrected likelihood ratio test in [21] are excluded from our comparison, since the former is similar to T_s and the latter is not defined for $p \geq n$, see [21]. Empirical sizes and powers of the studied tests are reported in two scenarios of distribution:

$$\text{(I) } w_{11} \sim N(0, 1), \qquad \text{(II) } w_{11} \sim \frac{\Gamma(5, 2) - 5/2}{\sqrt{5/4}}.$$

In the second scenario, w_{11} is actually a standardized Gamma-distributed random variable with $E(w_{11}^4) = 4.2$. The dimensional settings are $n = 50, 100, 150, 200, 250, 300$ and $p = cn$ with $c = 1, 5, 10, 20$. The nominal significance level is fixed at $\alpha = 0.05$, and the number of independent replications is 10 000.

Tables 1 and 2 collect empirical sizes of the three tests, where the covariance matrix is set to be $\Sigma_p = I_p$. The results show that, under the scenario (I) of normal distribution, all the sizes are close to the nominal significance level. Under the scenario (II) of Gamma distribution, there is slight size distortion for the tests when p and n are small, but this distortion fades away as p and n are increased.

To compare the powers of the three tests, we study two models under the alternative hypothesis.

Model 1. $\Sigma_p = I_p + \sqrt{c} \cdot \mathrm{diag}(\underbrace{1.2, \ldots, 1.2}_{n/50}, \underbrace{0, \ldots, 0}_{p-n/50})$,

Model 2. $\Sigma_p = I_p + \mathrm{diag}(\underbrace{0.5, \ldots, 0.5}_{p/2}, \underbrace{0, \ldots, 0}_{p/2})$.

Covariance matrices defined in Models 1 and 2 are both diagonal and near sphericity. The difference between them is that, in Model 1, there is only a small cluster of diagonal elements that deviate from the bulk, while in Model 2 half of the diagonal elements are apart from the rest.

Table 1
Empirical sizes in percentages of the tests T_s, T_f, and T_m for standard normal random variables.

n	$c = 1$			$c = 5$			$c = 10$			$c = 20$		
	T_s	T_f	T_m	T_s	T_f	T_m	T_s	T_f	T_m	T_s	T_f	T_m
50	5.53	5.08	5.63	5.12	5.03	5.67	4.88	5.08	5.25	4.75	5.14	5.31
100	4.93	5.47	5.70	5.41	5.52	5.68	4.96	4.83	4.84	4.94	4.83	5.08
150	5.14	5.12	5.34	5.15	5.61	5.56	5.10	4.95	5.17	5.25	5.01	5.23
200	5.00	5.47	5.35	5.33	5.23	5.61	5.09	5.16	5.17	5.15	4.82	5.07
250	5.09	5.31	5.40	4.79	5.53	5.38	4.97	4.93	5.21	5.38	5.26	5.40
300	4.74	5.07	5.03	5.25	5.19	5.31	4.91	5.34	5.13	5.22	4.90	5.10

Table 2
Empirical sizes in percentages of the tests T_s, T_f, and T_m for standardized Gamma(5, 2) random variables.

n	$c = 1$			$c = 5$			$c = 10$			$c = 20$		
	T_s	T_f	T_m	T_s	T_f	T_m	T_s	T_f	T_m	T_s	T_f	T_m
50	6.56	6.68	7.42	5.45	6.35	6.80	5.28	5.90	6.05	5.43	5.19	5.95
100	4.96	6.23	6.12	5.26	5.72	5.78	5.37	5.96	5.91	5.31	5.05	5.51
150	5.41	6.14	6.20	5.24	5.44	5.55	5.39	5.37	5.57	5.16	4.99	5.36
200	5.57	5.77	5.95	4.86	5.44	5.31	5.05	5.11	5.09	5.25	4.99	5.26
250	4.99	5.18	5.29	5.22	5.53	5.44	5.42	5.59	5.79	5.37	5.15	5.32
300	5.06	5.42	5.36	5.17	5.82	5.73	5.10	5.32	5.25	5.54	5.39	5.30

Empirical powers under Model 1 are plotted in Fig. 1 for normal variables and Fig. 2 for Gamma variables. It is clearly shown that all the powers grow to 1 as the dimensions increase, which demonstrates the consistency of the three tests. Moreover, the shift in data distribution does not seem to affect the powers. In comparison of performance, our proposed test T_m is more powerful than T_s and is comparable to T_f. Power results under Model 2 are illustrated in Fig. 3 for normal variables and Fig. 4 for Gamma variables, which exhibit a different pattern from that under Model 1. This time T_m is comparable to T_s and both of them dominant T_f in power. It seems that the test T_f fails to capture the difference between the null and the alternative hypotheses with the studied dimensions, especially when c is large. Synthesizing all the results, we may conclude that the proposed test is more robust than the other competitive ones.

5. Concluding remarks

This paper investigates the problem of testing for sphericity of covariance matrices in high-dimensions. Tests proposed in [18,8] are corrected to accommodate situations where the underlying distribution is not normal, based on which a new test procedure is developed by considering the maximum of the two modified statistics. Simulation studies claim that the new test is more robust in power than the original ones against various models.

The asymptotic results derived in this paper assume that the observations $\mathbf{x}_i$'s have zero mean vector. When this is not true we may replace the sample covariance matrix S_n with its centralized version S_n^*,

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

where $\bar{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i/n$. In such a situation, all the conclusions hold if we substitute $c_n^* := p/(n-1)$ for $c_n = p/n$ in the test statistics, see [21,22].

Acknowledgments

Weiming Li's research is supported by National Natural Science Foundation of China, No. 11401037, Fundamental Research Funds for the Central Universities, No. 2014RC0905, and FDCT078/2013/A3 of Macau.

Appendix

A.1. Proof of Lemma 1

Suppose that Assumptions (a)–(c) hold, from [17], the spectral distribution F_n converges weakly to a limiting distribution $F^{c,H}$, moreover the Stieltjes transform $s_n(z)$ of F_n converges almost surely to $s(z)$, the Stieltjes transform of $F^{c,H}$. Let F^{c_n,H_p} be

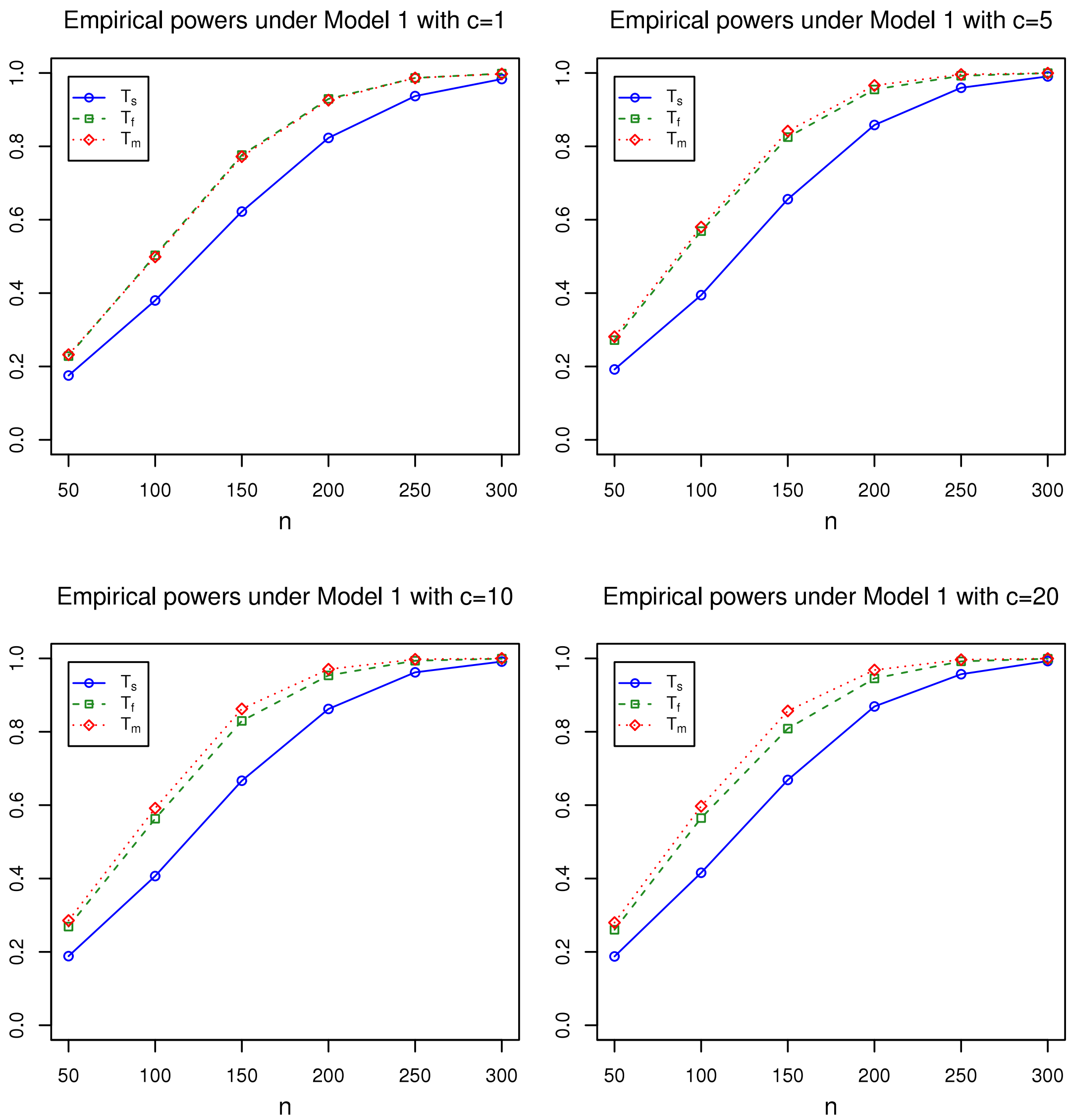

Fig. 1. Empirical powers of T_s, T_f, and T_m under Model 1 in the scenario of normal distribution. The significance level is $\alpha = 0.05$.

a distribution derived from $F^{c,H}$ by replacing c and H with c_n and H_p, respectively. Then the kth moments of F^{c_n,H_p} and $F^{c,H}$ are given by

$$\beta_k = \int t^k dF^{c_n,H_p}(t) \quad \text{and} \quad \tilde{\beta}_k = \int t^k dF^{c,H}(t),$$

$k = 1, 2, \ldots$. Obviously, $\beta_k \to \tilde{\beta}_k$ as $(n, p) \to \infty$. On the other hand, from [14], the relationship between α_k and β_k, $k = 1, 2, 3, 4$, are

$$\alpha_1 = \beta_1, \qquad \alpha_2 = \beta_2 - c_n\beta_1^2, \qquad \alpha_3 = \beta_3 - 3c_n\beta_2\beta_1 + 2c_n^2\beta_1^3,$$
$$\alpha_4 = \beta_4 - 4c_n\beta_3\beta_1 - 2c_n\beta_2^2 + 10c_n^2\beta_2\beta_1^2 - 5c_n^3\beta_1^4.$$

When the support of H is bounded, the support of $F^{c,H}$ is also bounded. From Lebesgue's dominated convergence theorem, for sufficient large (n, p),

$$\hat{\beta}_k = \int x^k dF_n(x) = -\frac{1}{2\pi i}\oint_{\mathcal{C}} z^k s_n(z)dz \xrightarrow{\text{a.s.}} -\frac{1}{2\pi i}\oint_{\mathcal{C}} z^k s(z)dz = \tilde{\beta}_k,$$

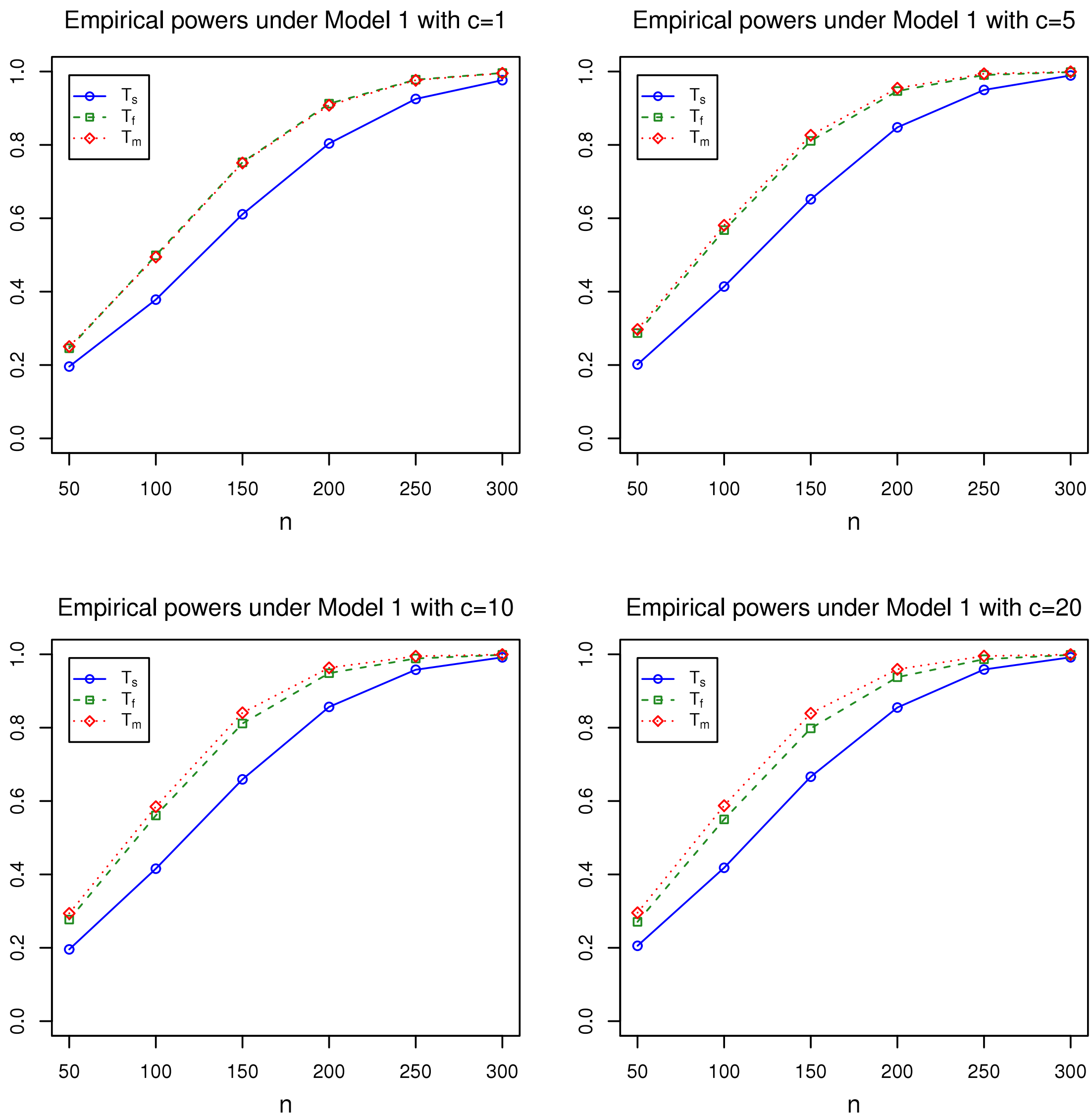

Fig. 2. Empirical powers of T_s, T_f, and T_m under Model 1 in the scenario of Gamma distribution. The significance level is $\alpha = 0.05$.

where the contour C is simple, closed, taken in the positive direction in the complex plane, and enclosing the support of $F^{c,H}$. Therefore, we get

$$\hat{\alpha}_k - \alpha_k \xrightarrow{\text{a.s.}} 0, \quad k = 1, 2, 3, 4,$$

as $(n, p) \to \infty$, which is the first conclusion of this lemma.

For the second conclusion, we first derive the limiting distribution of $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)'$. Applying the central limit theorem in [2] to functions $f(x) = x^k,\ k = 1, 2, 3, 4$, yields

$$n(\hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \hat{\beta}_3 - \beta_3, \hat{\beta}_4 - \beta_4)' \xrightarrow{D} N_4(\mathbf{m}, \Gamma),$$

where the mean vector $\mathbf{m} = (m_j)$ with

$$m_j = -\frac{1}{2\pi \mathrm{i}} \oint_{C_1} \frac{z^j \underline{s}^3(z) \int t^2 (1 + t\underline{s}(z))^{-3} dH(t)}{(1 - c\int \underline{s}^2(z) t^2 (1 + t\underline{s}(z))^{-2} dH(t))^2} dz, \tag{5}$$

and the covariance $\Gamma = (\gamma_{ij})$ with its entries

$$\gamma_{ij} = -\frac{1}{2\pi^2 c^2} \oint_{C_2} \oint_{C_1} \frac{z_1^i z_2^j}{(\underline{s}(z_1) - \underline{s}(z_2))^2} \underline{s}'(z_1) \underline{s}'(z_2) dz_1 dz_2, \tag{6}$$

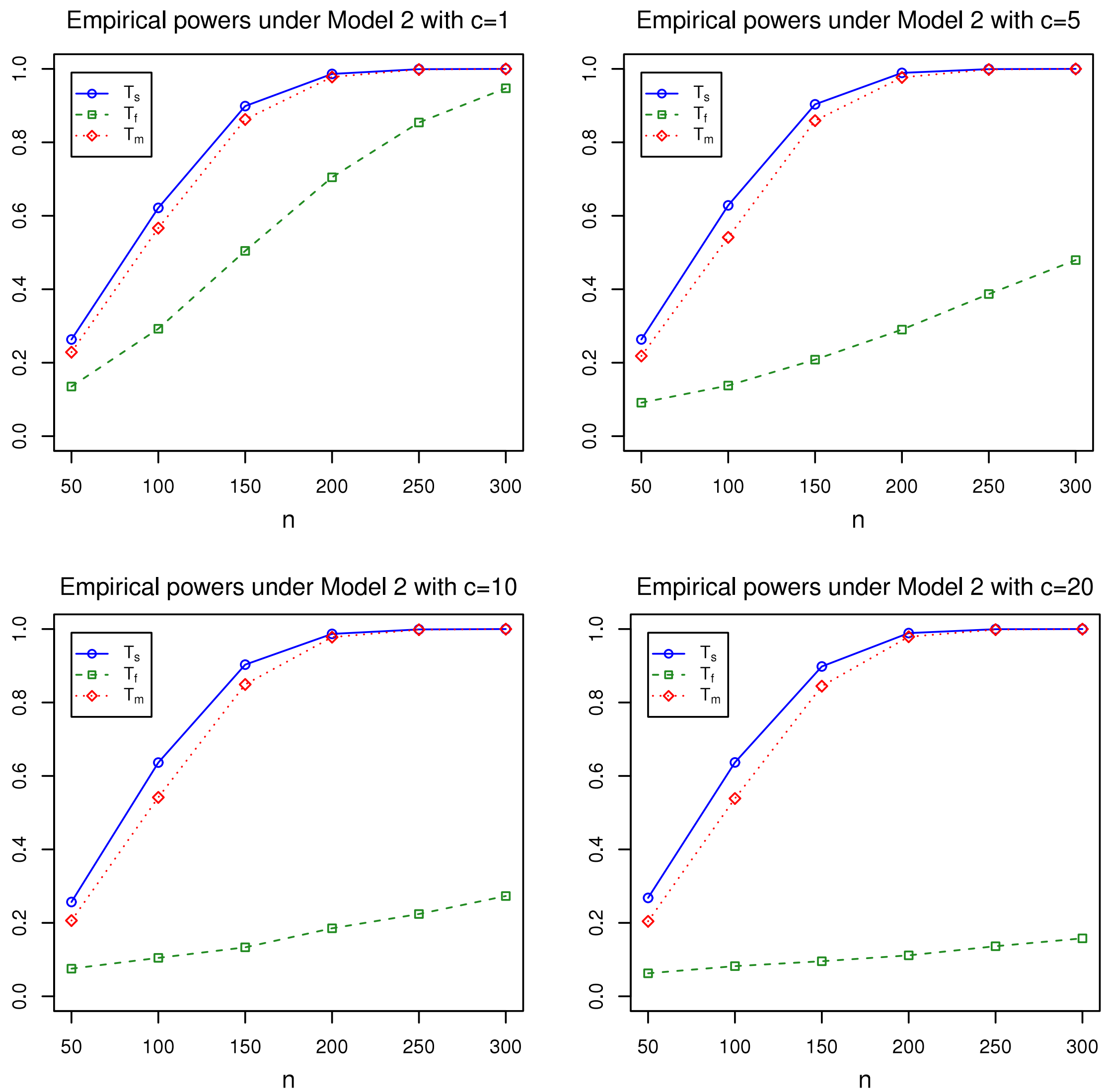

Fig. 3. Empirical powers of T_s, T_f, and T_m under Model 2 in the scenario of normal distribution. The significance level is $\alpha = 0.05$.

where $\underline{s}(z) = -(1-c)/z + cs(z)$ and the contours C_1 and C_2 in (5) and (6) are simple, closed, non-overlapping, taken in the positive direction in the complex plane, and each enclosing the support of $F^{c,H}$. The contour integrals in (5) and (6) have been figured out in [13], from which we can get

$$m_1 = 0, \qquad m_2 = \tilde{\alpha}_2, \qquad m_3 = 3(c\tilde{\alpha}_1\tilde{\alpha}_2 + \tilde{\alpha}_3),$$
$$m_4 = 6c^2\tilde{\alpha}_1^2\tilde{\alpha}_2 + 5c\tilde{\alpha}_2^2 + 12c\tilde{\alpha}_1\tilde{\alpha}_3 + 6\tilde{\alpha}_4,$$

and

$$\gamma_{ij} = \frac{2}{c^2}\sum_{l=0}^{i-1}(i-l)u_{i,l}u_{j,i+j-l},$$

where $u_{s,t}$ is the coefficient of z^t in the Taylor expansion of $(-1 - c\sum_{l=1}^{\infty}\tilde{\alpha}_l(-z)^l)^s$.

Next, we use the Delta method and Slutsky's theorem to get the final result. Let $\mathbf{u} = (x, y, z, w)'$ and define a vector function G_n,

$$G_n(\mathbf{u}) = \Bigg(x,\ \tau_2(y - c_n x^2),\ \tau_3(z - 3c_n xy + 2c_n^2 x^3),\ \tau_4\Bigg(w - 4c_n xz - \frac{2n^2+3n-6}{n^2+n+2}c_n y^2 + \frac{10n^2+12n}{n^2+n+2}c_n^2 x^2 y - \frac{5n^2+6n}{n^2+n+2}c_n^3 x^4\Bigg)\Bigg)'.$$

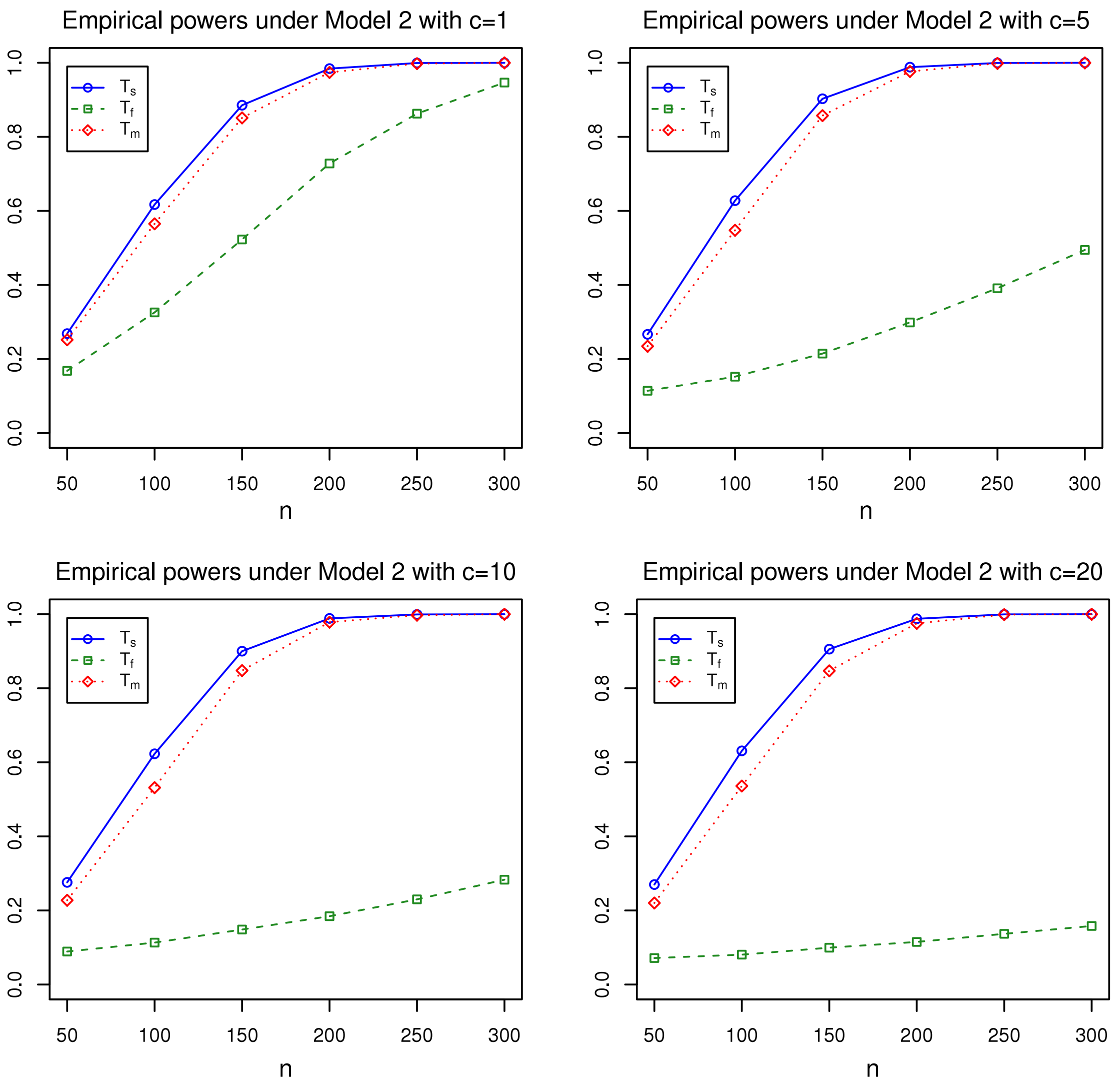

Fig. 4. Empirical powers of T_s, T_f, and T_m under Model 2 in the scenario of Gamma distribution. The significance level is $\alpha = 0.05$.

It is clear that G_n has continuous partial derivative at $\mathbf{b} := (\beta_1, \beta_2, \beta_3, \beta_4)'$ and the Jacobian matrix $J_n(\mathbf{b}) = \partial G_n/\partial \mathbf{u}|_{\mathbf{u}=\mathbf{b}}$ converges to a limit $J(\mathbf{b})$, as $(n, p) \to \infty$,

$$J(\mathbf{b}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2c\tilde{\alpha}_1 & 1 & 0 & 0 \\ 3c^2\tilde{\alpha}_1^2 - 3c\tilde{\alpha}_2 & -3c\tilde{\alpha}_1 & 1 & 0 \\ -4c^3\tilde{\alpha}_1^3 + 8c^2\tilde{\alpha}_1\tilde{\alpha}_2 - 4c\tilde{\alpha}_3 & 6c^2\tilde{\alpha}_1^2 - 4c\tilde{\alpha}_2 & -4c\tilde{\alpha}_1 & 1 \end{pmatrix}.$$

Therefore we get, as $(n, p) \to \infty$,

$$n\left(\hat{\alpha}_1 - \alpha_1, \hat{\alpha}_2 - \alpha_2, \hat{\alpha}_3 - \alpha_3, \hat{\alpha}_4 - \alpha_4\right)' + n\left((\alpha_1, \alpha_2, \alpha_3, \alpha_4)' - G_n(\mathbf{b})\right) \xrightarrow{D} N_4(J(\mathbf{b})\mathbf{m}, J(\mathbf{b})\Gamma J'(\mathbf{b})).$$

Elementary calculations reveal that

$$n((\alpha_1, \alpha_2, \alpha_3, \alpha_4)' - G_n(\mathbf{b})) \to J(\mathbf{b})\mathbf{m} = (0, \tilde{\alpha}_2, 3\tilde{\alpha}_3, 6\tilde{\alpha}_4 + c\tilde{\alpha}_2^2)',$$

and $J(\mathbf{b})\Gamma J'(\mathbf{b}) := A = (a_{ij})$ with its entries

$$a_{11} = 2\tilde{\alpha}_2/c, \qquad a_{12} = 4\tilde{\alpha}_3/c, \qquad a_{13} = 6\tilde{\alpha}_4/c, \qquad a_{14} = 8\tilde{\alpha}_5/c,$$

$$a_{22} = 4(2\tilde{\alpha}_4/c + \tilde{\alpha}_2^2), \qquad a_{23} = 12(\tilde{\alpha}_5/c + \tilde{\alpha}_3\tilde{\alpha}_2),$$

$$a_{24} = 8(2\tilde{\alpha}_6/c + 2\tilde{\alpha}_4\tilde{\alpha}_2 + \tilde{\alpha}_3^2), \qquad a_{33} = 6(3\tilde{\alpha}_6/c + 3\tilde{\alpha}_4\tilde{\alpha}_2 + 3\tilde{\alpha}_3^2 + c\tilde{\alpha}_2^3),$$

$$a_{34} = 24(\tilde{\alpha}_7/c + \tilde{\alpha}_5\tilde{\alpha}_2 + 2\tilde{\alpha}_4\tilde{\alpha}_3 + c\tilde{\alpha}_3\tilde{\alpha}_2^2),$$
$$a_{44} = 8(4\tilde{\alpha}_8/c + 6\tilde{\alpha}_4^2 + 4\tilde{\alpha}_6\tilde{\alpha}_2 + 8\tilde{\alpha}_5\tilde{\alpha}_3 + 4c\tilde{\alpha}_4\tilde{\alpha}_2^2 + 8c\tilde{\alpha}_3^2\tilde{\alpha}_2 + c^2\tilde{\alpha}_2^4),$$

which complete the proof.

A.2. Proof of Lemma 2

Under the assumptions of this lemma, the conditions of Theorem 1.4 in [16] are all satisfied, which implies

$$n(\hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \hat{\beta}_3 - \beta_3, \hat{\beta}_4 - \beta_4)' \xrightarrow{D} N_4(\bar{\mathbf{m}}, \bar{\Gamma}),$$

where the mean vector $\bar{\mathbf{m}} = (\bar{m}_j)$ with

$$\begin{aligned}\bar{m}_j &= m_j - \frac{E(w_{11}^4) - 3}{2\pi \mathrm{i}} \oint_{C_1} \frac{z^j \underline{s}^3(z) \int t^2(1 + t\underline{s}(z))^{-3} dH(t)}{1 - c\int \underline{s}^2(z) t^2 (1 + t\underline{s}(z))^{-2} dH(t)} dz \\ &:= m_j + \Delta I_j,\end{aligned} \tag{7}$$

and the covariance $\bar{\Gamma} = (\bar{\gamma}_{ij})$ with its entries

$$\begin{aligned}\bar{\gamma}_{ij} &= \gamma_{ij} - \frac{E(w_{11}^4) - 3}{4\pi^2 c} \oint_{C_2}\oint_{C_1} z_1^i z_2^j \frac{d^2}{dz_1 dz_2} \int \frac{t^2 \underline{s}(z_1)\underline{s}(z_2) dH(t)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} dz_1 dz_2 \\ &:= \gamma_{ij} + \Delta J_{ij},\end{aligned} \tag{8}$$

where $\Delta = E(w_{11}^4) - 3$, m_j and γ_{ij} are defined in (5) and (6), respectively.

Without loss of generality, let the contour C_2 enclose C_1 and both of them be away from the support S_F of $F^{c,H}$. Denote the image of C_i under $\underline{s}(z)$ be

$$\underline{s}(C_i) = \{\underline{s}(z) : z \in C_i\}, \quad i = 1, 2.$$

Then, following similar arguments in [13], the two contours $\underline{s}(C_1)$ and $\underline{s}(C_2)$ are also simple, closed, and non-overlapping, and are taken in the negative direction. Moreover, $\underline{s}(C_2)$ encloses $\underline{s}(C_1)$ and both of them enclose zero. Let

$$P(\underline{s}) = -1 + c\int \frac{t\underline{s}}{1 + t\underline{s}} dH(t) \quad \text{and} \quad Q(\underline{s}) = \int \frac{t^2}{(1 + t\underline{s})^3} dH(t),$$

the integral I_j in (7) becomes

$$\begin{aligned}I_j &= -\frac{1}{2\pi \mathrm{i}} \oint_{C_1} z^j \underline{s}(z)\underline{s}'(z) \int t^2(1 + t\underline{s}(z))^{-3} dH(t) dz \\ &= -\frac{1}{2\pi \mathrm{i}} \oint_{\underline{s}(C_1)} \frac{P^j(\underline{s}) Q(\underline{s})}{\underline{s}^{j-1}} d\underline{s} \\ &= \begin{cases} 0, & j = 1, \\ \dfrac{1}{(j-2)!} \left[P^j(z) Q(z)\right]^{(j-2)} \Big|_{z=0}, & j \geq 2, \end{cases}\end{aligned}$$

while the integral J_{ij} in (8) is

$$\begin{aligned}J_{ij} &= -\frac{1}{4\pi^2 c} \oint_{C_2}\oint_{C_1} z_1^i z_2^j \int \frac{t^2 \underline{s}'(z_1)\underline{s}'(z_2) dH(t)}{(1 + t\underline{s}(z_1))^2 (1 + t\underline{s}(z_2))^2} dz_1 dz_2 \\ &= -\frac{1}{4\pi^2 c} \oint_{\underline{s}(C_2)}\oint_{\underline{s}(C_1)} \frac{P(\underline{s}_1)^i P(\underline{s}_2)^j}{\underline{s}_1^i \underline{s}_2^j} \int \frac{t^2 dH(t)}{(1 + t\underline{s}_1)^2 (1 + t\underline{s}_2)^2} d\underline{s}_1 d\underline{s}_2 \\ &= \frac{1}{c} \int \frac{t^2}{(i-1)!(j-1)!} \left[\frac{P^i(z)}{(1 + tz)^2}\right]^{(i-1)} \Bigg|_{z=0} \left[\frac{P^j(z)}{(1 + tz)^2}\right]^{(j-1)} \Bigg|_{z=0} dH(t),\end{aligned}$$

where the contour integrals are calculated from the Cauchy integral theorem. Following similar arguments in the proof of Lemma 1, we may get

$$n(\hat{\alpha}_1 - \alpha_1, \hat{\alpha}_2 - \alpha_2, \hat{\alpha}_3 - \alpha_3, \hat{\alpha}_4 - \alpha_4)' \xrightarrow{D} N_4(J(\mathbf{b})(\bar{\mathbf{m}} - \mathbf{m}), J(\mathbf{b})\bar{\Gamma}J'(\mathbf{b})).$$

Elementary calculations reveal that

$$J(\mathbf{b})(\bar{\mathbf{m}} - \mathbf{m}) = \Delta \cdot (0, \tilde{\alpha}_2, 3\tilde{\alpha}_3, 6\tilde{\alpha}_4)'$$

and

$$J(\mathbf{b})(\bar{\Gamma}-\Gamma)J'(\mathbf{b}) = \frac{\Delta}{c}\cdot\begin{pmatrix}\tilde{\alpha}_2 & 2\tilde{\alpha}_3 & 3\tilde{\alpha}_4 & 4\tilde{\alpha}_5\\ 2\tilde{\alpha}_3 & 4\tilde{\alpha}_4 & 6\tilde{\alpha}_5 & 8\tilde{\alpha}_6\\ 3\tilde{\alpha}_4 & 6\tilde{\alpha}_5 & 9\tilde{\alpha}_6 & 12\tilde{\alpha}_7\\ 4\tilde{\alpha}_5 & 8\tilde{\alpha}_6 & 12\tilde{\alpha}_7 & 16\tilde{\alpha}_8\end{pmatrix}.$$

The proof is then completed.

A.3. Proof of Theorems 1

This theorem follows from Lemmas 1 and 2 by a standard application of the Delta method, we thus only calculate the limiting mean vector μ and covariance matrix Ω. Let $\mathbf{u} = (x, y, z, w)'$ and define a vector function G,

$$G(\mathbf{u}) = \left(\frac{y}{x^2}, \frac{w}{y^2}\right)'.$$

It is clear that G has continuous partial derivative at $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ and thus the Jacobian matrix $K_n(\mathbf{a}) = \partial G/\partial \mathbf{u}|_{\mathbf{u}=\mathbf{a}} \to K(\mathbf{a})$,

$$K(\mathbf{a}) = \begin{pmatrix} -\dfrac{2\tilde{\alpha}_2}{\tilde{\alpha}_1^3} & \dfrac{1}{\tilde{\alpha}_1^2} & 0 & 0\\ 0 & -\dfrac{2\tilde{\alpha}_4}{\tilde{\alpha}_2^3} & 0 & \dfrac{1}{\tilde{\alpha}_2^2}\end{pmatrix}.$$

If Σ_p is diagonal for all p large then, from Lemma 2, the liming mean vector μ is

$$\mu = \Delta \cdot K(\mathbf{a})(0, \tilde{\alpha}_2, 3\tilde{\alpha}_3, 6\tilde{\alpha}_4)' = \Delta\cdot(\tilde{\alpha}_2/\tilde{\alpha}_1^2, 4\tilde{\alpha}_4/\tilde{\alpha}_2^2)'$$

and the covariance matrix is $\Omega = K(\mathbf{a})BK'(\mathbf{a})$ where the matrix B is defined in (2). If $E(w_{11}^3) = 3$ then, from Lemma 1, we get $\mu = 0$ and $\Omega = K(\mathbf{a})AK'(\mathbf{a})$ where the matrix A is defined in (1).

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., John Wiley & Sons, Hoboken, NJ, 2003.
[2] Z.D. Bai, J.W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices, Ann. Probab. 32 (2004) 553–605.
[3] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., Springer, New York, 2010.
[4] Q. Berthet, P. Rigollet, Optimal detection of sparse principal components in high dimension, Ann. Statist. 41 (4) (2013) 1780–1815.
[5] T.T. Cai, Z.M. Ma, Optimal hypothesis testing for high dimensional covariance matrices, Bernoulli 19 (5B) (2013) 2359–2388.
[6] S.X. Chen, L.X. Zhang, P.S. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (490) (2010) 810–819.
[7] T.J. Fisher, On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size, J. Statist. Plann. Inference 142 (1) (2012) 312–326.
[8] T.J. Fisher, X.Q. Sun, C.M. Gallagher, A new test for sphericity of the covariance matrix for high dimensional data, J. Multivariate Anal. 101 (10) (2010) 2554–2570.
[9] S. John, Some optimal multivariate tests, Biometrika 58 (1971) 123–127.
[10] S. John, The distribution of a statistic used for testing sphericity of normal distributions, Biometrika 59 (1972) 169–173.
[11] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2) (2001) 295–327.
[12] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (4) (2002) 1081–1102.
[13] W.M. Li, Y.L. Qin, Bias-reduced estimators of moments of a population spectral distribution and their applications. Mauscript, 2015.
[14] A. Nica, R. Speicher, Lectures on the Combinatorics of Free Probability, Cambridge University Press, New York, 2006.
[15] A. Onatski, M.J. Moreira, M. Hallin, Asymptotic power of sphericity tests for high-dimensional data, Ann. Statist. 41 (3) (2013) 1204–1231.
[16] G.M. Pan, W. Zhou, Central limit theorem for signal-to-interference ratio of reduced rank linear receiver, Ann. Appl. Probab. 18 (2008) 1232–1270.
[17] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.
[18] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2) (2005) 251–272.
[19] M.S. Srivastava, Some tests criteria for the covariance matrix with fewer observations than the dimension, Acta Comment. Univ. Tartu. Math. (10) (2006) 77–93.
[20] M.S. Srivastava, T. Kollo, D. von Rosen, Some tests for the covariance matrix with fewer observations than the dimension under non-normality, J. Multivariate Anal. 102 (6) (2011) 1090–1103.
[21] Q.W. Wang, J.F. Yao, On the sphericity test with large-dimensional observations, Electron. J. Stat. 7 (2013) 2164–2192.
[22] S.R. Zheng, Z.D. Bai, A note on central limit theorems for linear spectral statistics of large dimensional F Matrix, 2013. arXiv:1305.0339.