Asymptotic distribution of the LR statistic for equality of the smallest eigenvalues in high-dimensional principal component analysis

Yasunori Fujikoshi, Takayuki Yamada*, Daisuke Watanabe, Takakazu Sugiyama

Department of Mathematics, Graduate School of Science and Engineering, Chuo University, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Received 23 May 2006
Available online 22 December 2006

Abstract

This paper deals with the distribution of the LR statistic for testing the hypothesis that the smallest eigenvalues of a covariance matrix are equal. We derive an asymptotic null distribution of the LR statistic when the dimension \( p \) and the sample size \( N \) approach infinity, while the ratio \( p/N \) converging on a finite nonzero limit \( c \in (0,1) \). Numerical simulations revealed that our approximation is more accurate than the classical chi-square-type approximation as \( p \) increases in value.

© 2006 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: 62H10; 62E20

Keywords: Asymptotic distribution; High-dimensional principal component; LR statistic; Equality of the smallest eigenvalues

1. Introduction

Let \( x_1, \ldots, x_N \) be a sample of size \( N \) from a \( p \)-variate normal distribution with an unknown mean vector and an unknown covariance matrix \( \Sigma \). Let \( \lambda_1 \geq \cdots \geq \lambda_p > 0 \) be the eigenvalues of \( \Sigma \). We consider the problem of testing the hypothesis that the smallest \( m = p - q \) eigenvalues of \( \Sigma \) are equal; that is, we test the following hypothesis:

\[
H_0: \lambda_{q+1} = \cdots = \lambda_p (= \lambda, \text{unknown}).
\]

* Corresponding author.
E-mail address: yma@math.chuo-u.ac.jp (T. Yamada).
This problem is important, because in applications of principal component analysis, we are usually interested in the number of $q$ for which the first $q$ principal components include most of the information contained within the covariance matrix $\Sigma$ and the remaining principal components have almost the same small variances.

Let

$$\bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_i$$

and

$$S = \frac{1}{N-1} \sum_{j=1}^{N} (x_j - \bar{x})(x_j - \bar{x})'.$$

The likelihood ratio test rejects the hypothesis $H_0$ for small values of

$$V = \frac{\prod_{j=q+1}^{p} \phi_j(S)}{\left\{ \frac{1}{m} \sum_{j=q+1}^{p} \phi_j(S) \right\}^m},$$

where for any square matrix $A$ the notation $\phi_j(A)$ denotes its $j$th largest eigenvalue (see e.g., [4, p. 406]). Note that the likelihood ratio statistic $V$ is defined only for $N - 1 \geq p$. When we consider the null distribution of $V$, there is no loss of generality that $\Sigma$ may be diagonal, because the eigenvalues of $S$ are the same as those of $Q'SQ$ for any orthogonal matrix $Q$. Furthermore, $V$ is invariant due to the multiplication of $S$ by a positive scalar. Therefore, under $H_0$, we may assume that $\Sigma = \text{diag}(\lambda_1/\lambda, \ldots, \lambda_q/\lambda, 1, \ldots, 1)$. For notational simplicity, we write

$$\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_q, 1, \ldots, 1).$$

(1)

Here, $\lambda_i$ should read as $\tilde{\lambda}_i = \lambda_i/\lambda$.

In the special case $q = 0$, Wakaki [8] derived the limiting distribution. However, Schott [7] proposed a statistic that can be used for $n = N - q - 1 < p$ and derived the asymptotic null distribution of this statistic when $p/n$ tends to some positive constant. By applying the aforementioned ideas to our case, we derive an asymptotic null distribution of the LR statistic in a high-dimensional framework such that

$A1 : q; \text{fix}, \ n \rightarrow \infty, \ m \rightarrow \infty, \ m/n \rightarrow c \in (0, 1)$.

$A2 : \lambda_j = O(m), \ j = 1, \ldots, q.$

We used numerical simulations to demonstrate that our approximation is more accurate than the classical chi-square-type approximation as $p$ increases in value.

2. Main theorem

Let $S$ be partitioned as

$$\begin{pmatrix}
S_{11} & S_{12} \\
S_{12}' & S_{22}
\end{pmatrix},$$

where $S_{11}, S_{12}$ and $S_{22}$ are $q \times q, q \times m$ and $m \times m$ matrices, respectively. First, we consider

$$\tilde{V} = \frac{\det (S_{22:1})}{\left\{ \frac{1}{m} \text{tr} (S_{22:1}) \right\}^m},$$
instead of $V$, where $S_{22,1} = S_{22} - S_{12}^{-1}S_{11}$. Without loss of generality from the invariance property of the statistic $\tilde{V}$, $S_{22,1}$ may be distributed as a central Wishart distribution $W_m(n, I_m)$. The characteristic function of log $\tilde{V}$ is expressed as

$$m^{mit} \frac{\Gamma(mn/2) \Gamma_m(n/2)}{\Gamma(mn/2 + mit) \Gamma_m(n/2)}$$

where the function $\Gamma_m(\cdot)$ is the $m$-variate gamma function (see e.g., [4, pp. 342 and 62]) defined as

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{j=1}^m \Gamma(a - \frac{1}{2}(j-1)).$$

Therefore, the cumulant-generating function of log $\tilde{V}$ is expanded as follows:

$$\log E[\exp(it \log \tilde{V})] = it \mu_{m,n} + \frac{1}{2}(it)^2 \sigma_{m,n}^2 + \frac{1}{6}(it)^3 \gamma_{3,m,n} + \cdots,$$

where

$$\mu_{m,n} = m \log m - m \psi\left(\frac{mn}{2}\right) + \psi_m\left(\frac{n}{2}\right),$$

$$\sigma_{m,n}^2 = \psi'\left(\frac{n}{2}\right) - m^2 \psi'\left(\frac{mn}{2}\right),$$

$$\gamma_{k,m,n} = \psi_m^{(k-1)}\left(\frac{n}{2}\right) - m^k \psi^{(k-1)}\left(\frac{mn}{2}\right)$$

for any integer $k \geq 3$. Here, $\psi(\cdot)$ is the digamma function defined by

$$\psi(a) = \frac{d}{da} \log \Gamma(a) = -C + \sum_{k=0}^{\infty} \left( \frac{1}{1 + k} - \frac{1}{a + k} \right) = O(\log a),$$

where $C$ is the Euler constant, and

$$\psi_m(a) = \sum_{j=1}^m \psi\left(a - \frac{1}{2}(j-1)\right).$$

Noting that

$$\psi^{(s)}(a) = \sum_{k=0}^{\infty} \frac{-s(-1)^s}{(a+k)^{s+1}} = O(a^{-s}),$$

we can see that

$$\sigma_{m,n}^2 = O(1) \quad \text{and} \quad \gamma_{k,m,n} = O_p(n^{-(k-2)})$$

under the condition A1. Therefore, the characteristic function of $(\log \tilde{V} - \mu_{m,n})/\sigma_{m,n}$ can be expanded as

$$\exp \left\{ \frac{1}{2} (it)^2 \left[ 1 + \frac{1}{6\sigma_{m,n}^2} (it)^3 \gamma_{3,m,n} \right] + O(n^{-2}) \right\}.$$

Formally inverting this function leads to the following theorem.
Theorem 1. Under conditions (1) and A1, the distribution function of \( \log \tilde{V} \) can be expanded as

\[
\Pr \left( \frac{\log \tilde{V} - \mu_{m,n}}{\sigma_{m,n}} \leq x \right) = \Phi(x) - \frac{1}{6\sigma_{m,n}^3} \gamma_{3,m,n} \phi(x) (x^2 - 1) + O(n^{-2}),
\]

where \( \Phi(\cdot) \) and \( \phi(\cdot) \) are the distribution function of the standard normal distribution and its derivative. The constants \( \mu_{m,n}, \sigma_{m,n} \) and \( \gamma_{3,m,n} \) are defined by (2) with \( m = p - q \) and \( n = N - q - 1 \).

Note that Theorem 1 was essentially derived by Wakaki [8], since he has presented a similar asymptotic expansion of the null distribution of the LR criterion for testing \( \Sigma = \sigma^2 I_p \).

Next, we link \( \log V \) and \( \log \tilde{V} \), which is crucial for our derivation of an asymptotic distribution of \( \log V \).

Lemma 1. Under conditions (1), A1 and A2, \( \log V \) converges in probability to \( \log \tilde{V} \).

The proof of Lemma 1 is presented in Section 4. We can write

\[
\frac{\log V - \mu_{m,n}}{\sigma_{m,n}} = \frac{\log \tilde{V} - \mu_{m,n}}{\sigma_{m,n}} + \frac{\log V - \log \tilde{V}}{\sigma_{m,n}}.
\]

According to Theorem 1 and Lemma 1, the first term on the right-hand side of the above equation converges in distribution on \( N(0, 1) \), and the second term converges in probability to 0. Therefore, using Slutzky’s theorem (see e.g., [6]) we have the following main theorem (Theorem 2).

Theorem 2. Under conditions (1), A1 and A2, it holds that

\[
\frac{\log V - \mu_{m,n}}{\sigma_{m,n}} \overset{d}{\to} N(0, 1),
\]

where \( \overset{d}{\to} \) denotes convergence in distribution and the constants \( \mu_{m,n} \) and \( \sigma_{m,n} \) are defined by (2) with \( m = p - q \) and \( n = N - q - 1 \).

3. Simulation results

This section presents the results of numerical simulations to demonstrate the effectiveness of our theorem of the asymptotic normality of \( Z_{m,n} = (\log V - \mu_{m,n})/\sigma_{m,n} \) for some values of \( m \) and \( n \). In all our simulations, we took the common smallest eigenvalue to be 1 and we set \( \lambda_i = \rho_i m / (1 - \sum_{j=1}^q \rho_j) \), which is the same model as in Schott [7], for \( i = 1, \ldots, q \), \( N = 100 \), \( p = 10, 20, 30, 40, 50, 60, 70, 80 \) and 90, and \( q = 2 \). In Table 1, we list the estimated significance levels for \( Z_{m,n} \) calculated by using 1,000,000 repetitions for the case in which \( \rho_1 = 0.56 \) and \( \rho_2 = 0.24 \) with nominal significance levels of 0.01, 0.05, 0.50, 0.95, and 0.99.

To compare these results with the test based on the classical chi-square approximation (see e.g., [2]), we list the significance levels for \( -c_{m,n} \log \Lambda \) in Table 2 using the same setting as for the simulation presented in Table 1, where \( c_{m,n} = 1 - (2m^2 - m + 2)/(6mn) \) is Bartlett correction factor (see e.g., [2]).

Tables 1 and 2 illustrate that our new approximation is more accurate than the classical chi-square approximation, except when \( p = 10 \). Furthermore, the data in Table 2 indicate that the
chi-square approximation becomes increasingly inaccurate as the value of $p$ increases for a fixed value of $N$.

4. Proof of Lemma 1

In this section, we prove Lemma 1 under conditions (1), A1 and A2. For this purpose, it is sufficient to show that

(i) $\log \det (S_{22,1}) - \log \left\{ \prod_{j=q+1}^{p} \phi_j(S) \right\} = o_p(1),$

(ii) $m \log \left\{ \frac{1}{m} \text{tr} S_{22,1} \right\} - m \log \left\{ \frac{1}{m} \sum_{j=q+1}^{p} \phi_j(S) \right\} = o_p(1).$

Previously, Schott [7] proved that

$$\frac{1}{m} \sum_{j=q+1}^{p} \phi_j(S) - \frac{1}{m} \text{tr} S_{22,1} = o_p(n^{-1}). \quad (3)$$

Note that $S$ and $S_{22,1}$ are theoretically positive definite matrix under condition A1.

Let the inverse matrix of $S$ can be expressed as

$$S^{-1} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$
where $T_{22}$ is expressed as $S_{22}^{-1}$. Applying the Poincaré separation theorem (see e.g., [5]) to $S^{-1}$, we find that $\phi_j(T_{22}) \leq \phi_j(S^{-1})$ for $j = 1, \ldots, m$, and so that $\phi_{m-j+1}(T_{22}^{-1}) \leq \phi_{m-j+1}(S)$. Hence, it holds that

$$1 \leq \prod_{j=1}^m \frac{\phi_{m-j+1}(T_{22}^{-1})}{\phi_{p-j+1}(S)} = \frac{\det(S_{22}^{-1})}{\prod_{j=q+1}^p \phi_j(S)},$$

or equivalently

$$0 \leq \log \det(S_{22}^{-1}) - \log \left( \prod_{j=q+1}^p \phi_j(S) \right) = \sum_{j=1}^m \left( \log \phi_j(S_{22}^{-1}) - \log \phi_{q+j}(S) \right).$$

Let $F_j(t)$ be a real-valued function on $[0, 1]$ defined by

$$F_j(t) = \log \{(\phi_j(S_{22}^{-1}) + 1) - t\} - \log \{(\phi_{q+j}(S) + 1) - t\}.$$

It is found from Tayler’s theorem that there exists $\theta_j$ in $(0, 1)$ such that $F_j(1) = F_j(0) + F'_j(\theta_j)$, and hence

$$\log \phi_j(S_{22}^{-1}) - \log \phi_{q+j}(S) = \log (\phi_j(S_{22}^{-1}) + 1) - \log (\phi_{q+j}(S) + 1)$$

$$+ \frac{\phi_j(S_{22}^{-1}) - \phi_{q+j}(S)}{(\phi_j(S_{22}^{-1}) + 1 - \theta_j)(\phi_{q+j}(S) + 1 - \theta_j)}.$$

Since $\log (1 + x) - \log (1 + y) \leq \log (1 + x - y)$ for $0 < y < x$, it holds that

$$\log (\phi_j(S_{22}^{-1}) + 1) - \log (\phi_{q+j}(S) + 1)$$

$$\leq \log (1 + \phi_j(S_{22}^{-1}) - \phi_{q+j}(S))$$

$$\leq (\phi_j(S_{22}^{-1}) - \phi_{q+j}(S)),$$

for $j = 1, \ldots, m$, where the last inequality follows from the fact that $\log (1 + x) \leq x$ for $x \geq 0$. Furthermore, for $j = 1, \ldots, m$, we have

$$\frac{\phi_j(S_{22}^{-1}) - \phi_{q+j}(S)}{(\phi_j(S_{22}^{-1}) + 1 - \theta_j)(\phi_{q+j}(S) + 1 - \theta_j)} \leq \frac{\phi_j(S_{22}^{-1}) - \phi_{q+j}(S)}{(\phi_j(S_{22}^{-1}) + 1 - \theta_j)(\phi_{q+j}(S) + 1 - \theta_j)} \leq \frac{\phi_j^2(S)}{\phi_{q+j}^2(S)}.$$

Therefore, it holds that

$$0 \leq \sum_{j=1}^m \{ \log \phi_j(S_{22}^{-1}) - \log \phi_{q+j}(S) \}$$

$$\leq \left( 1 + \frac{1}{\phi_p^2(S)} \right) \sum_{j=1}^m \{ \phi_j(S_{22}^{-1}) - \phi_{q+j}(S) \}.$$

It follows from Bai and Yin [1] that

$$\phi_p(S) \rightarrow (1 - c^{1/2})^2 \text{ a.s., since } p/n \rightarrow c \in (0, 1).$$

(5)
This implies that the right-hand side of the inequality (4) is \( o_p(1) \), since result (3) holds. Consequently, result (i) holds.

Otherwise, result (3) implies that
\[
1 \leq \frac{\text{tr}(S_{22,1})}{\sum_{j=q+1}^{p} \phi_j(S)} \leq 1 + o_p(n^{-1})
\]
since \( \sum_{j=q+1}^{p} \phi_j(S) \geq m \phi_p(S) \) and result (5). Therefore, we can find that
\[
0 \leq \log \{\text{tr}(S_{22,1})\} - \log \left\{ \sum_{j=q+1}^{p} \phi_j(S) \right\} = o_p(n^{-1}),
\]
which leads to the result (ii). From (i) and (ii), it is clear that \( \log V \) converges in probability to \( \log \tilde{V} \).

Note that Lemma 1 can be derived under the condition represented by a general subsequence for \( m \) and \( n \) as in Lodoit and Wolf [3] and Schott [7].

**Acknowledgments**

The authors would like to thank two referees for their useful comments and careful readings.

**References**