Note on asymptotic null distributions of LR statistics for testing covariance matrix under growth curve model when the number of the observation points is large

Takayuki Yamada

(Received March 29, 2011; Revised May 12, 2012)

Abstract. This paper is concerned with the testing problem about the covariance matrix under growth curve model. The testing problems treated in this paper are the following problems: (i) the problem of testing that a covariance matrix is equal to a specified positive definite matrix, (ii) the problem of testing for sphericity of the covariance matrix, and (iii) the problem of intraclass model for the covariance matrix. We give asymptotic distributions of the null distributions for the likelihood ratio statistics under an asymptotic framework that the total sample size n go to infinity, the number of the observation points p go to infinity, and p/n go to a constant $c \in (0,1)$. Simulation reveals that the proposed approximations have good accuracies compared with the classical chi-square approximations.

 $AMS\ 2010\ Mathematics\ Subject\ Classification.$ primary 62H10 ; secondary 62H15.

Key words and phrases. Growth curve model, likelihood ratio test, testing problem for a specified covariance matrix, sphericity, intraclass model, high-dimensional approximation.

§1. Introduction

Let Y be a $p \times N$ observation matrix whose column vectors are independently distributed. A growth curve model for Y can be expressed as

$$(1.1) Y = B\Theta A + \mathcal{E},$$

where \boldsymbol{B} is a known $p \times q$ matrix of constants with the rank q, which is referred to within individual design matrix; Θ is $q \times k$ parameter matrix; \boldsymbol{A}

is a known $k \times N$ matrix of constants with the rank k, which is referred to between individual design matrix; \mathcal{E} is an $p \times N$ error matrix. Assume that $\mathcal{E} \sim N_{p,N}(\mathbf{O}, \mathbf{\Sigma} \otimes \mathbf{I}_N)$, matrix variate normal distribution with mean \mathbf{O} and the covariance matrix of $\text{vec}(\mathcal{E}')$ is $\mathbf{\Sigma} \otimes \mathbf{I}_N$, where the notation " \otimes " denotes Kronecker's product. This paper is concerned with testing problem for $\mathbf{\Sigma}$. We consider the following null hypotheses:

- (i) $H_1: \Sigma = G$, where G is a specified positive definite matrix.
- (ii) $H_2: \Sigma = \lambda G$, where λ is unknown parameter.
- (iii) $H_3: \Sigma = \Sigma_{\rm I} = \lambda_1 G + \lambda_2 w w'$, where λ_1 and λ_2 are unknown parameters, and w is a p vector of constants which lies in the column space of B.

Likelihood ratio(LR) statistics and the moments have been obtained by Srivastava and Khatri [4] and Khatri [2]. Edgeworth expansions of these statistics for large n can be obtained. It, however, gets worth when p is moderate.

In this paper, we derive asymptotic approximations for the null distributions of the likelihood ratio statistics for these three test under asymptotic framework A1:

A1:
$$n \to \infty$$
, $p \to \infty$, $\frac{p}{n} \to c \in (0,1)$.

The paper is organized as follows. In Section 2, we propose asymptotic null distributions of LR statistics for testing the null hypotheses (i), (ii) and (iii) under A1. In Section 3, we do small scale simulation to confirm the precision of the proposed approximations of the null distributions.

§2. Asymptotic null distribution of LR statistics when n and p go to infinity together

For testing the null hypothesis $H_1: \Sigma = \Sigma_0$ under (1.1), Srivastava and Khatri [4] proposed modified likelihood ratio statistic

$$\Lambda_1 = \left(\frac{2e}{n}\right)^{pn/2} |\boldsymbol{G}^{-1}\{\boldsymbol{S} + (\boldsymbol{I}_p - \boldsymbol{T}\boldsymbol{S}^{-1})\boldsymbol{S}_1(\boldsymbol{I}_p - \boldsymbol{T}\boldsymbol{S}^{-1})'\}|^{n/2}$$
$$\cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{G}^{-1}\left[\boldsymbol{S} + \left\{\boldsymbol{G} - \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{G}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}'\right\}\boldsymbol{G}^{-1}\boldsymbol{S}_1\right]\right),$$

where

$$S = Y\{I_N - A'(AA')^{-1}A\}Y,$$

 $T = B(B'S^{-1}B)^{-1}B',$
 $S_1 = YA'(AA')^{-1}AY,$
 $n = N - k.$

Let

$$V_1 = \left\{ \left(\frac{e}{n} \right)^{-pn/2} \Lambda_1 \right\}^{2/n}.$$

The h-th moment of V_1 is given by Srivastava and Khatri [4] as

$$E[V_1^h] = 2^{ph} \left(1 + \frac{2}{n} h \right)^{-np/2 - k(p-q)/2 - ph} \prod_{j=1}^q \frac{\Gamma\left(\frac{n-p+j}{2} + h\right)}{\Gamma\left(\frac{n-p+j}{2}\right)}$$

$$\cdot \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2} + h\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)}.$$

Hence the characteristic function $C_{V_1}(t)$ of $-\log V_1$ is given by

$$C_{V_1}(t) = 2^{-pit} \left(1 - \frac{2}{n}it \right)^{-np/2 - k(p-q)/2 + pit} \prod_{j=1}^{q} \frac{\Gamma\left(\frac{n-p+j}{2} - it\right)}{\Gamma\left(\frac{n-p+j}{2}\right)}$$

$$\cdot \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2} - it\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)}.$$

Taylor's series expansion for $K_1(t) = \log C_{V_1}(t)$ can be expressed as

$$K_1(t) = \mu_1(it) + \frac{(it)^2}{2!}\sigma_1^2 + \sum_{s=3}^{\infty} \frac{(it)^s}{s!}\kappa_1^{(s)},$$

where

$$\mu_{1} = -p \log 2 + \frac{np + k(p - q)}{n} - \sum_{j=1}^{q} \psi\left(\frac{n - p + j}{2}\right)$$

$$- \sum_{j=1}^{p-q} \psi\left(\frac{n + k - p + q + j}{2}\right),$$

$$\sigma_{1}^{2} = 2\left\{-p + \frac{np + k(p - q)}{2n}\right\} \left(\frac{2}{n}\right) + \sum_{j=1}^{q} \psi'\left(\frac{n - p + j}{2}\right)$$

$$+ \sum_{j=1}^{p-q} \psi'\left(\frac{n + k - p + q + j}{2}\right),$$

40

and

$$\kappa_1^{(s)} = s! \left\{ -\frac{p}{s-1} + \frac{np + k(p-q)}{sn} \right\} \left(\frac{2}{n} \right)^{s-1} + (-1)^s \left(\sum_{j=1}^q \psi^{(s-1)} \left(\frac{n-p+j}{2} \right) + \sum_{j=1}^{p-q} \psi^{(s-1)} \left(\frac{n+k-p+q+j}{2} \right) \right).$$

Here, $\psi^{(s-1)}(.)$ stands for polygamma function defined by

(2.1)
$$\psi^{(s-1)}(a) = \frac{d^{s+1}}{da^{s+1}} \log \Gamma(a)$$

$$= \begin{cases}
-C + \sum_{k=0}^{\infty} \left(\frac{1}{1+k} - \frac{1}{k+a}\right) & (s=0) \\
\sum_{k=0}^{\infty} \frac{(-1)^{s+1} s!}{(k+a)^{s+1}} & (s=1,2,\cdots),
\end{cases}$$

where C is the Euler constant. Let

(2.2)
$$Z_1 = \frac{-\log V_1 - \mu_1}{\sigma_1}.$$

The characteristic function $C_{Z_1}(t)$ of Z_1 can be expanded as

$$C_{Z_1}(t) = \exp(-t^2/2) \left[1 + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \left\{ \sum_{s=3}^{\infty} \frac{(it)^s}{s!} \frac{\kappa_1^{(s)}}{\sigma_1^s} \right\}^{\ell} \right].$$

From (2.1),

$$\sigma_1^2 = O_0^*, \quad \kappa_1^{(s)} = O_{s-2}^* \quad (s = 3, 4, \cdots)$$

under asymptotic framework A1, where O_j^* denotes a term of j-th order with respect to (n^{-1}, p^{-1}) . Thus, Z_1 converges in distribution to the standard normal distribution under asymptotic framework A1.

For testing the null hypothesis $H_2: \Sigma = \lambda G$ under (1.1), Khatri [2] proposed likelihood ratio statistic

$$\Lambda_2 = \left(\frac{|S||I_p + (S^{-1} - S^{-1}B(B'S^{-1}B)^{-1}B'S^{-1})S_1|}{|G|[\{\operatorname{tr} G^{-1}S + \operatorname{tr}(G^{-1} - G^{-1}B(B'G^{-1}B)^{-1}B'G^{-1})S_1\}/p]^p}\right)^{N/2}.$$

Let

$$V_2 = \Lambda_2^{2/N}.$$

The h-th moment of V_2 has been given in Khatri [2]. But his result has a mistake, so should be improved as

$$E[V_2^h] = p^{ph} \prod_{j=1}^q \frac{\Gamma\left(\frac{n-p+j}{2} + h\right)}{\Gamma\left(\frac{n-p+j}{2}\right)} \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2} + h\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)} \frac{\Gamma\left(\frac{np+k(p-q)}{2}\right)}{\Gamma\left(\frac{np+k(p-q)}{2} + ph\right)}.$$

Hence the characteristic function $C_{V_2}(t)$ of $-\log V_2$ is given by

$$C_{V_2}(t) = p^{pit} \prod_{j=1}^q \frac{\Gamma\left(\frac{n-p+j}{2} - it\right)}{\Gamma\left(\frac{n-p+j}{2}\right)} \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2} - it\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)} \frac{\Gamma\left(\frac{np+k(p-q)}{2}\right)}{\Gamma\left(\frac{np+k(p-q)}{2} - pit\right)}.$$

Taylor's series expansion for $K_2(t) = \log C_{V_2}(t)$ can be expressed as

$$K_2(t) = \mu_2(it) + \frac{(it)^2}{2!}\sigma_2^2 + \sum_{s=3}^{\infty} \frac{(it)^s}{s!}\kappa_2^{(s)},$$

where

$$\begin{split} \mu_2 &= -p \log p - \sum_{j=1}^q \psi\left(\frac{n-p+j}{2}\right) - \sum_{j=1}^{p-q} \psi\left(\frac{n+k-p+q+j}{2}\right) \\ &+ p \psi\left(\frac{np+k(p-q)}{2}\right), \\ \sigma_2^2 &= \sum_{j=1}^q \psi'\left(\frac{n-p+j}{2}\right) + \sum_{j=1}^{p-q} \psi'\left(\frac{n+k-p+q+j}{2}\right) - p^2 \psi'\left(\frac{np+k(p-q)}{2}\right), \\ \kappa_2^{(s)} &= (-1)^s \left\{\sum_{j=1}^q \psi^{(s-1)}\left(\frac{n-p+j}{2}\right) + \sum_{j=1}^{p-q} \psi^{(s-1)}\left(\frac{n+k-p+q+j}{2}\right) - p^s \psi^{(s-1)}\left(\frac{np+k(p-q)}{2}\right)\right\}. \end{split}$$

Let

(2.3)
$$Z_2 = \frac{-\log V_2 - \mu_2}{\sigma_2}.$$

The characteristic function $C_{Z_2}(t)$ of Z_2 can be expanded as

$$C_{Z_2}(t) = \exp(-t^2/2) \left[1 + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \left\{ \sum_{s=3}^{\infty} \frac{(it)^s}{s!} \frac{\kappa_2^{(s)}}{\sigma_2^s} \right\}^{\ell} \right].$$

From (2.1),

$$\sigma_2^2 = O_0^*, \quad \kappa_2^{(s)} = O_{s-2}^* \quad (s = 3, 4, \cdots)$$

under asymptotic framework A1. Thus, Z_2 converges in distribution to the standard normal distribution under asymptotic framework A1.

For testing the null hypothesis $H_3: \Sigma = \Sigma_{\rm I}$ under (1.1), Khatri [2] proposed likelihood ratio statistic

$$\Lambda_3 = \left(\frac{|{\bm S}||{\bm I}_p + ({\bm S}^{-1} - {\bm S}^{-1}{\bm B}({\bm B}'{\bm S}^{-1}{\bm B})^{-1}{\bm B}'{\bm S}^{-1}){\bm S}_1|}{n^p|{\bm G}|\hat{\lambda}_1^{p-1}(\hat{\lambda}_1 + {\bm w}'{\bm G}^{-1}{\bm w}\hat{\lambda}_2)}\right)^{N/2},$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are solutions of the following equalities:

$$n(p-1)\hat{\lambda}_1 = \operatorname{tr} \mathbf{G}^{-1} \mathbf{S} + \operatorname{tr} (\mathbf{G}^{-1} - \mathbf{G}^{-1} \mathbf{B} (\mathbf{B}' \mathbf{G}^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{G}^{-1}) \mathbf{S}_1$$
$$- \frac{\mathbf{w}' \mathbf{G}^{-1} \mathbf{S} \mathbf{G}^{-1} \mathbf{w}}{\mathbf{w}' \mathbf{G}^{-1} \mathbf{w}}.$$
$$n(\hat{\lambda}_1 + \mathbf{w}' \mathbf{G}^{-1} \mathbf{w} \hat{\lambda}_2) = \frac{\mathbf{w}' \mathbf{G}^{-1} \mathbf{S} \mathbf{G}^{-1} \mathbf{w}}{\mathbf{w}' \mathbf{G}^{-1} \mathbf{w}}.$$

Let

$$V_3 = \Lambda_3^{2/N}.$$

The h-th moment of V_3 has been given in Khatri [2]. But his result has a mistake, so should be improved as

$$\begin{split} E[V_3^h] = & \frac{(p-1)^{(p-1)h}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n(p-1)+k(p-q)}{2}\right)}{\Gamma\left(\frac{n}{2}+h\right)\Gamma\left(\frac{n(p-1)+k(p-q)}{2}+(p-1)h\right)} \prod_{j=1}^q \frac{\Gamma\left(\frac{n-p+j}{2}+h\right)}{\Gamma\left(\frac{n-p+j}{2}\right)} \\ & \times \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2}+h\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)}. \end{split}$$

Hence the characteristic function $C_{V_3}(t)$ of $-\log V_3$ is given by

$$\begin{split} C_{V_3}(t) = & \frac{(p-1)^{-(p-1)it} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n(p-1)+k(p-q)}{2}\right)}{\Gamma\left(\frac{n}{2}-it\right) \Gamma\left(\frac{n(p-1)+k(p-q)}{2}-(p-1)it\right)} \prod_{j=1}^q \frac{\Gamma\left(\frac{n-p+j}{2}-it\right)}{\Gamma\left(\frac{n-p+j}{2}\right)} \\ & \times \prod_{j=1}^{p-q} \frac{\Gamma\left(\frac{n+k-p+q+j}{2}-it\right)}{\Gamma\left(\frac{n+k-p+q+j}{2}\right)}. \end{split}$$

Taylor's series expansion for $K_3(t) = \log C_{V_3}(t)$ can be expressed as

$$K_3(t) = \mu_3(it) + \frac{(it)^2}{2!}\sigma_3^2 + \sum_{s=3}^{\infty} \frac{(it)^s}{s!}\kappa_3^{(s)},$$

where

$$\mu_{3} = -(p-1)\log(p-1) + \psi\left(\frac{n}{2}\right) - \sum_{j=1}^{q} \psi\left(\frac{n-p+j}{2}\right)$$

$$-\sum_{j=1}^{p-q} \psi\left(\frac{n+k-p+q+j}{2}\right)$$

$$+(p-1)\psi\left(\frac{n(p-1)+k(p-q)}{2}\right),$$

$$\sigma_{3}^{2} = -\left\{\psi'\left(\frac{n}{2}\right) - \sum_{j=1}^{q} \psi'\left(\frac{n-p+j}{2}\right) - \sum_{j=1}^{p-q} \psi'\left(\frac{n+k-p+q+j}{2}\right)\right\}$$

$$+(p-1)^{2}\psi'\left(\frac{n(p-1)+k(p-q)}{2}\right)\right\},$$

$$\kappa_{3}^{(s)} = (-1)^{s}\left\{\psi^{(s-1)}\left(\frac{n}{2}\right) - \sum_{j=1}^{q} \psi^{(s-1)}\left(\frac{n-p+j}{2}\right)\right\}$$

$$-\sum_{j=1}^{p-q} \psi^{(s-1)}\left(\frac{n+k-p+q+j}{2}\right)$$

$$+(p-1)^{s}\psi^{(s-1)}\left(\frac{n(p-1)+k(p-q)}{2}\right)\right\}.$$

Let

(2.4)
$$Z_3 = \frac{-\log V_3 - \mu_3}{\sigma_3}.$$

The characteristic function $C_{\mathbb{Z}_3}(t)$ of \mathbb{Z}_3 can be expanded as

$$C_{Z_3}(t) = \exp(-t^2/2) \left[1 + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \left\{ \sum_{s=3}^{\infty} \frac{(it)^s}{s!} \frac{\kappa_3^{(s)}}{\sigma_3^s} \right\}^{\ell} \right].$$

From (2.1),

$$\sigma_3^2 = O_0^*, \quad \kappa_3^{(s)} = O_{s-2}^* \quad (s = 3, 4, \cdots)$$

under asymptotic framework A1. Thus, Z_3 converges in distribution to the standard normal distribution under asymptotic framework A1.

Theorem 1. Assume that the null hypotheses H_i , i = 1, 2, 3, are true. Under asymptotic framework A1, Z_i defined in (2.2), (2.3) and (2.4) converge in distribution to the standard normal distributions, respectively.

Rigorous proofs of theorems can be obtained by using the same method written in Wakaki [5] or Kato et al. [1].

§3. Simulation study

In this section, we did the small scale simulation to confirm the precision of the approximation proposed in Theorem 1. Data was generated by $Y = B\Theta A + \mathcal{E}$, where

$$\mathbf{B'} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1/(p-1) & 2/(p-1) & \cdots & (p-1)/(p-1) \\ 0 & \{1/(p-1)\}^2 & \{2/(p-1)\}^2 & \cdots & \{(p-1)/(p-1)\}^2 \end{pmatrix},
\mathbf{\Theta} = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 7 & 2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{1'}_{N_1} & \mathbf{0'}_{N_2} \\ \mathbf{0'}_{N_1} & \mathbf{1'}_{N_2} \end{pmatrix},$$

 $\mathcal{E} \sim N_{p,N}(\boldsymbol{O}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_N)$. We treated the case in which $\boldsymbol{G} = \boldsymbol{I}_p$, $\boldsymbol{w} = \boldsymbol{1}_p$, $\lambda = 1$ and $\lambda_1 = \lambda_2 = 1/2$. Here, $\boldsymbol{1}_p$ denotes the *p*-dimensional vector whose elements are all 1. For simplicity, we studied the case in which $N_1 = N_2$. The number of observation *p* varied over the values 10,30,50,70 and 90. We calculated the actual probability for nominal from 1,000,000 simulations when the nominal level is 0.05. Actual probabilities is almost monotone decreasing as *p*, and larger than 0.05. A classical approximate chi square tests reject the null hypothesis H_i if $-2\tau_i \log W_i$ is larger than the upper α point of chi square distribution with f_i degrees of freedom for i = 1, 2, 3, where

$$\begin{split} W_1 &= \Lambda_1, \quad W_2 = \Lambda_2^{n/N}, \quad W_3 = \Lambda_3^{n/N}, \\ f_1 &= \frac{p(p+1)}{2}, \quad f_2 = \frac{(p+2)(p-1)}{2}, \quad f_3 = \frac{p(p+1)}{2} - 2, \\ \tau_1 &= 1 - \frac{1}{n} \left\{ \frac{2p^2 + 3p - 1}{6(p+1)} - \frac{k(p-q)(p-q-k+1)}{p(p+1)} \right\}, \\ \tau_2 &= 1 - \frac{1}{n} \left\{ \frac{2p^2 + p + 2}{6p} - \frac{k(p-q)(p^2 + p + pq - kq - 2)}{p(p-1)(p+2)} \right\}, \\ \tau_3 &= 1 - \frac{1}{n} \left\{ \frac{p(p+1)(2p-3)}{6(p-1)(p^2 + p + 4)} - \frac{k(p-q)(p^2 - pq + k + q - kq - 3)}{(p-1)(p^2 + p + 4)} \right\}. \end{split}$$

Here, τ_i is the Bartlett collection factor. We calculated the actual probability for nominal when the nominal level is 0.05. From Table 1 we can see that actual probability is almost monotone decreasing as p, and larger than 0.05. On the other hand, actual probability is monotone increasing as p, and gets close to 1 when p is 90. Consequently, approximations based on Theorem 1 have good accuracies.

p	(i)		(ii)		(iii)	
	Z_1	$-2\tau_1 \log W_1$	Z_2	$-2\tau_2 \log W_2$	Z_3	$-2\tau_3 \log W_3$
10	0.0597	0.0509	0.0595	0.0558	0.0593	0.0517
30	0.0536	0.0635	0.0536	0.0717	0.0540	0.0659
50	0.0523	0.1547	0.0522	0.1720	0.0523	0.1588
70	0.0521	0.6928	0.0518	0.7208	0.0520	0.7014
90	0.0522	1.0000	0.0520	1.0000	0.0519	1.0000

Table 1: Comparison of actual probabilities of Z_i and $-2\tau_i \log W_i$, i = 1, 2, 3, for N = 100 when the nominal level is 0.05.

§4. Conclusion

This paper deals with the likelihood ratio test for testing some structures of the covariance matrix under the growth curve model proposed by Potthoff and Roy [3]. We derive the asymptotic null distributions under asymptotic framework that the sample size N and the dimension p go to infinity together with p < N. Simulation results show that our proposed tests have good accuracies when p is relatively large compared to N, which reveal that classical chi-square tests have bad. We remark that error bound for our proposed approximation shall be derived along with Wakaki [5] or Kato et al. [1].

References

- [1] N. Kato, T. Yamada and Y. Fujikoshi, *High-dimensional asymptotic expansion* of LR statistic for testing intraclass correlation structure ant its error bound, J. Multivariate Anal., **101** (2009), 101–112.
- [2] C. G. Khatri, Testing some covariance structures under a growth curve model, J. Multivariate Anal., 3 (1973), 102–116.
- [3] R. F. Potthoff and S. N. Roy, A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika, **51** (1964), 313-326.
- [4] M. S. Srivastava and C. G. Khatri, An Introduction to Multivariate Statistics, North-Holland, New York, 1979.
- [5] H. Wakaki, Error bounds for high-dimensional Edgeworth expansions for some tests on covariance matrices, in Proceeding of International Conference on Multivariate Statistical Modeling and High Dimensional Data Mining held in Kayseri on June 19–23, 2008.

Takayuki Yamada Risk Analysis Research Center, The Institute of Statistical Mathematics 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan E-mail: yma801228@gmail.com