

A Note on Approximation of Likelihood Ratio Statistic in Exploratory Factor Analysis

Masanori Ichikawa

Graduate School of Global Studies, Tokyo University of Foreign Studies, Tokyo, Japan

Email: ichikawa.m@tufs.ac.jp

Received 11 September 2015; accepted 27 October 2015; published 30 October 2015

Copyright © 2015 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In normal theory exploratory factor analysis, likelihood ratio (LR) statistic plays an important role in evaluating the goodness-of-fit of the model. In this paper, we derive an approximation of the LR statistic. The approximation is then used to show explicitly that the expectation of the LR statistic agrees with the degrees of freedom of the asymptotic chi-square distribution.

Keywords

Factor Analysis, Likelihood Ratio Statistic, Maximum Likelihood Estimation

1. Introduction

Factor analysis [1] [2] is used in various fields to study interdependence among a set of observed variables by postulating underlying factors. We consider the model of exploratory factor analysis in the form

$$\Sigma = \Lambda\Lambda' + \Psi, \quad (1)$$

where Σ is the $p \times p$ covariance matrix of observed variables, Λ is a $p \times m$ matrix of factor loadings, and Ψ is a diagonal matrix of error variances with $\psi_{ii} > 0$ ($i = 1, \dots, p$). Under the assumption of multivariate normal distributions for observations, the parameters are estimated with the method of maximum likelihood and the goodness-of-fit of the model can be judged by using the likelihood ratio (LR) test for testing the null hypothesis $\Sigma = \Lambda\Lambda' + \Psi$ for a specified m against the alternative that Σ is unconstrained. From the theory of LR tests, the degrees of freedom, ν , of the asymptotic chi-square distribution is the difference between the number of free parameters on the alternative model and the null model. In (1), Σ remains unchanged if Λ is replaced by $\Lambda\mathbf{T}$ for any $m \times m$ orthogonal matrix \mathbf{T} . Hence, $m(m-1)/2$ restrictions are required to eliminate this indeterminacy. Then, the difference between the number of nonduplicated elements in Σ and the number of free parameters in Λ and Ψ is given by

$$v = \frac{p(p+1)}{2} - \left\{ p(m+1) + \frac{m(m-1)}{2} \right\} = \frac{1}{2} \{ (p-m)^2 - (p+m) \}. \quad (2)$$

2. LR Statistic in Exploratory Factor Analysis

2.1. Approximation of LR Statistic

Let \mathbf{S} be the usual unbiased estimator of Σ based on a random sample of size $N = n+1$ from the multivariate normal population $N_p(\mu, \Sigma)$ with $\Sigma = \Lambda\Lambda' + \Psi$. For the existence of consistent estimators, we assume that the solution Ψ of $\Sigma = \Lambda\Lambda' + \Psi$ is unique. A necessary condition for the uniqueness of the solution Ψ up to multiplication on the right of Λ by an orthogonal matrix is that each column of $\Lambda\mathbf{T}$ has at least three non-zero elements for every non-singular matrix \mathbf{T} ([3], Theorem 5.6). This condition implies that $\text{rank}(\Lambda) = m$.

The maximum Wishart likelihood estimators $\hat{\Lambda}$ and $\hat{\Psi}$ are defined as the values of Λ and Ψ that minimize

$$F(\Sigma) = \text{tr}(\mathbf{S}\Sigma^{-1}) - \log|\mathbf{S}\Sigma^{-1}| - p. \quad (3)$$

Then, $\hat{\Lambda}$ and $\hat{\Psi}$ can be shown to be the solutions of the following equations:

$$(\mathbf{S} - \hat{\Sigma})\hat{\Psi}^{-1}\hat{\Lambda} = \mathbf{0}, \quad (4)$$

$$\text{diag}(\mathbf{S} - \hat{\Sigma}) = \mathbf{0}, \quad (5)$$

where $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$. The motivation behind the minimization of $F(\Sigma)$ in (3) is that

$$\text{LR} = nF(\hat{\Sigma}), \quad (6)$$

that is, n times the minimum value $F(\hat{\Sigma})$ is the LR statistic described in the previous section. Under (4) and (5), $\text{tr}\mathbf{S}\hat{\Sigma}^{-1} = p$ and $(\mathbf{S} - \hat{\Sigma})\hat{\Sigma}^{-1} = (\mathbf{S} - \hat{\Sigma})\hat{\Psi}^{-1}$ can be shown to hold. Hence,

$$nF(\hat{\Sigma}) = -n \log|\mathbf{S}\hat{\Sigma}^{-1}| = -n \log|\mathbf{I}_p + (\mathbf{S} - \hat{\Sigma})\hat{\Sigma}^{-1}| = -n \log|\mathbf{I}_p + (\mathbf{S} - \hat{\Sigma})\hat{\Psi}^{-1}|.$$

From the second-order Taylor formula, we have an approximation of the LR statistic as

$$\text{LR} \approx \frac{1}{2} \text{tr} \left\{ n^{1/2} (\mathbf{S} - \hat{\Sigma})\hat{\Psi}^{-1} \right\}^2 = \sum_{i < j} \frac{(s_{ij} - \hat{\sigma}_{ij})^2}{\hat{\psi}_{ii}\hat{\psi}_{jj}}, \quad (7)$$

by virtue of (5) [1] [2]. While the approximation on the right hand side of (7) shows how the LR statistic is related to the sum of squares of standardized residuals [4], it does not enable us to investigate the distributional properties of the LR statistic. To overcome this difficulty, we express the LR statistic as a function of $\mathbf{U} = n^{1/2}(\mathbf{S} - \Sigma)$.

Let $\tilde{\Psi}$ and $\tilde{\Lambda}$ denote the terms of $n^{1/2}(\hat{\Psi} - \Psi)$ and $n^{1/2}(\hat{\Lambda} - \Lambda)$ linear in the elements of \mathbf{U} . Then we have the following proposition.

Proposition 1. *An approximation of the LR statistic is given by*

$$\frac{1}{2} \text{tr} \Phi \mathbf{U} \Phi \mathbf{U} - \frac{1}{2} \text{tr} \Phi \mathbf{U} \Phi \tilde{\Psi}, \quad (8)$$

where $\Phi = (\phi_{ij})$ is defined by

$$\Phi = \Psi^{-1} - \Psi^{-1} \Lambda \Gamma^{-1} \Lambda' \Psi^{-1}, \quad (9)$$

with $\Gamma = \Lambda' \Psi^{-1} \Lambda$.

Proof. By substituting $\mathbf{U}, \tilde{\Lambda}$, and $\tilde{\Psi}$ into (4) and (5) and considering only linear terms, we have

$$(\mathbf{U} - \tilde{\Sigma})\Psi^{-1}\Lambda = \mathbf{0}, \quad (10)$$

$$\text{diag}(\mathbf{U} - \tilde{\Sigma}) = \mathbf{0}, \quad (11)$$

where $\tilde{\Sigma} = \Lambda \tilde{\Lambda}' + \tilde{\Lambda} \Lambda' + \tilde{\Psi}$. From (10) we derive

$$\begin{aligned}\tilde{\Lambda} \Gamma &= (\mathbf{U} - \tilde{\Psi}) \Psi^{-1} \Lambda - \Lambda \tilde{\Lambda}' \Psi^{-1} \Lambda, \\ \tilde{\Lambda} \Lambda' &= (\mathbf{U} - \tilde{\Psi}) \Pi \Psi - \Lambda \tilde{\Lambda}' \Pi \Psi,\end{aligned}$$

where $\Pi = \Psi^{-1} \Lambda \Gamma^{-1} \Lambda' \Psi^{-1}$. Then

$$\begin{aligned}\Lambda \tilde{\Lambda}' &= \Psi \Pi (\mathbf{U} - \tilde{\Psi}) - \Psi \Pi \tilde{\Lambda} \Lambda' \\ &= \Psi \Pi (\mathbf{U} - \tilde{\Psi}) - \Psi \Pi (\mathbf{U} - \tilde{\Psi}) \Pi \Psi + \Psi \Pi \Lambda \tilde{\Lambda}' \Pi \Psi \\ &= \Psi \Pi (\mathbf{U} - \tilde{\Psi}) - \Psi \Pi (\mathbf{U} - \tilde{\Psi}) \Pi \Psi + \Lambda \tilde{\Lambda}' \Pi \Psi,\end{aligned}$$

by virtue of $\Psi \Pi \Lambda = \Lambda$. Thus,

$$\begin{aligned}\mathbf{U} - \tilde{\Sigma} &= \mathbf{U} - \tilde{\Psi} - \Lambda \tilde{\Lambda}' - \tilde{\Lambda} \Lambda' \\ &= (\mathbf{U} - \tilde{\Psi}) - (\mathbf{U} - \tilde{\Psi}) \Pi \Psi - \Psi \Pi (\mathbf{U} - \tilde{\Psi}) + \Psi \Pi (\mathbf{U} - \tilde{\Psi}) \Pi \Psi \\ &= (\mathbf{I}_p - \Psi \Pi) (\mathbf{U} - \tilde{\Psi}) (\mathbf{I}_p - \Pi \Psi) \\ &= \Psi \Phi (\mathbf{U} - \tilde{\Psi}) \Phi \Psi.\end{aligned}\tag{12}$$

By replacing $n^{1/2} (\mathbf{S} - \hat{\Sigma}) \hat{\Psi}^{-1}$ in (7) with $(\mathbf{U} - \tilde{\Sigma}) \Psi^{-1} = \Psi \Phi (\mathbf{U} - \tilde{\Psi}) \Phi$, we have

$$\frac{1}{2} \text{tr} \Phi (\mathbf{U} - \tilde{\Psi}) \Phi (\mathbf{U} - \tilde{\Psi}),$$

since $\Phi \Psi \Phi = \Phi$. It follows from (11) and (12) that

$$\text{diag} \Phi (\mathbf{U} - \tilde{\Psi}) \Phi = \mathbf{0},\tag{13}$$

thus establishing the desired result.

2.2. Evaluating Expectation

For the purpose of demonstrating the usefulness of the derived approximation, we show explicitly that the expectation of (8) agrees with the degrees of freedom, ν , in (2) of the asymptotic chi-square distribution. We now evaluate the expectation of (8) by using

$$\mathbb{E} u_{ab} u_{cd} = \sigma_{ac} \sigma_{bd} + \sigma_{ad} \sigma_{bc},\tag{14}$$

see, for example, Theorem 3.4.4 of [1]. By noting $\text{tr} \Phi \Sigma = \text{tr} \Phi \Psi = p - m$, we see that the expectation of the first term in (8) is

$$\begin{aligned}\frac{1}{2} \mathbb{E} \text{tr} \Phi \mathbf{U} \Phi \mathbf{U} &= \frac{1}{2} \sum_{a,b,c,d} \phi_{ad} \phi_{bc} \mathbb{E} u_{ab} u_{cd} = \frac{1}{2} \sum_{a,b,c,d} \phi_{ad} \phi_{bc} (\sigma_{ac} \sigma_{bd} + \sigma_{ad} \sigma_{bc}) \\ &= \frac{1}{2} \left\{ \text{tr} \Phi \Sigma + (\text{tr} \Phi \Sigma)^2 \right\} = \frac{1}{2} \left\{ (p - m) + (p - m)^2 \right\}.\end{aligned}\tag{15}$$

To evaluate the expectation of the second term in (8), we need to express $\tilde{\Psi}$ in terms of \mathbf{U} . Let the symbol \odot denote the Hadamard product of matrices, and define $\Xi = (\xi_{ij})$ by $\Xi = \Phi \odot \Phi$. Because Φ is positive semidefinite, so is Ξ [5]. If Ξ is positive definite, then (13) can be solved for $\tilde{\Psi}$ in terms of \mathbf{U} [3]. An expression of $\tilde{\psi}_{ii}$ is

$$\tilde{\psi}_{ii} = \text{tr} \Phi \Theta_i \Phi \mathbf{U},\tag{16}$$

where Θ_i is a diagonal matrix whose diagonal elements are the i -th column (row) of $\Xi^{-1} = (\xi^{ij})$ [6]. An interesting property of $\Phi \Theta_i \Phi$ is

$$[\Phi \Theta_i \Phi]_{jj} = \sum_g \phi_{jg} \xi^{gi} \phi_{gj} = \sum_g \xi_{jg} \xi^{gi} = \delta_{ij},\tag{17}$$

where δ_{ij} is the Kronecker delta with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Hence, we have

$$\begin{aligned}
 \frac{1}{2} \text{Etr} \mathbf{\Phi} \mathbf{U} \mathbf{\Phi} \tilde{\mathbf{\Psi}} &= \frac{1}{2} \sum_{i,a,b} \phi_{ia} \phi_{ib} \text{E} u_{ab} (\text{tr} \mathbf{\Phi} \mathbf{\Theta}_i \mathbf{\Phi} \mathbf{U}) \\
 &= \frac{1}{2} \sum_{i,a,b,c,d} \phi_{ia} \phi_{ib} [\mathbf{\Phi} \mathbf{\Theta}_i \mathbf{\Phi}]_{cd} \text{E} u_{ab} u_{cd} \\
 &= \frac{1}{2} \sum_{i,a,b,c,d} \phi_{ia} \phi_{ib} [\mathbf{\Phi} \mathbf{\Theta}_i \mathbf{\Phi}]_{cd} (\sigma_{ac} \sigma_{bd} + \sigma_{ad} \sigma_{bc}) \\
 &= \sum_i [\mathbf{\Phi} \mathbf{\Sigma} \mathbf{\Phi} \mathbf{\Theta}_i \mathbf{\Phi} \mathbf{\Sigma} \mathbf{\Phi}]_{ii} = \sum_i [\mathbf{\Phi} \mathbf{\Theta}_i \mathbf{\Phi}]_{ii} = p.
 \end{aligned} \tag{18}$$

By combining (15) and (18), we obtain the desired result.

References

- [1] Anderson, T.W. (2003) An Introduction to Multivariate Statistical Analysis. 3rd Edition, John Wiley & Sons, Hoboken.
- [2] Lawley, D.N. and Maxwell, A.E. (1971) Factor Analysis as a Statistical Method. 2nd Edition, Butterworths, London.
- [3] Anderson, T.W. and Rubin, H. (1956) Statistical Inference in Factor Analysis. In: Neyman, J., Ed., *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, Berkeley, 111-150.
- [4] Browne, M.W., MacCallum, R.C., Kim, C.-T., Andersen, B.L. and Gleser, R. (2002) When Fit Indices and Residuals are Incompatible. *Psychological Methods*, **7**, 403-421. <http://dx.doi.org/10.1037/1082-989x.7.4.403>
- [5] Horn, R.A. and Johnson, C.R. (1991) Topics in Matrix Analysis. Cambridge University Press, Cambridge. <http://dx.doi.org/10.1017/CBO9780511840371>
- [6] Ichikawa, M. and Konishi, S. (2002) Asymptotic Expansions and Bootstrap Approximations in Factor Analysis. *Journal of Multivariate Analysis*, **81**, 47-66. <http://dx.doi.org/10.1006/jmva.2001.1991>