Journal of Multivariate Analysis 149 (2016) 144–159

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Bias correction of the Akaike information criterion in factor analysis

Haruhiko Ogasawara

Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan

ARTICLE INFO

Article history:
Received 18 November 2015
Available online 25 April 2016

AMS subject classifications:
62E20
62F12
62H25

Keywords:
Asymptotic bias
AIC
Covariance structure analysis
Exploratory factor analysis
Non-normality

ABSTRACT

The higher-order asymptotic bias for the Akaike information criterion (AIC) in factor analysis or covariance structure analysis is obtained when the parameter estimators are given by the Wishart maximum likelihood. Since the formula of the exact higher-order bias is complicated, simple approximations which do not include unknown parameter values are obtained. Numerical examples with simulations show that the approximations are reasonably similar to their corresponding exact asymptotic values and simulated values. Simulations for model selection give consistently improved results by the approximate correction of the higher-order bias for the AIC over the usual AIC.

1. Introduction

Information criteria are used primarily for model selection. One of the typical information criteria is the Akaike information criterion (AIC, [1]). The AIC is given by the familiar formula of −2 times the sample log likelihood plus 2 times the number of parameters in a statistical model. The latter quantity is a correction term corresponding to the negative asymptotic bias of order $O(1)$ for the former main term as an estimator of −2 times the expected log predictive likelihood, which is an unknown criterion for model selection. Since the bias is negative, the positive correction term is often interpreted as a penalty for a relatively complicated model in model selection.

The popularity of the AIC is due to its simplicity of the fixed correction term independent of unknown population values of parameters. The Takeuchi information criterion (TIC, [44]) gives an alternative correction term corresponding to that of the AIC under possible model (distribution) misspecification. The TIC was also derived by Stone [41] in the context of cross validation. For the TIC and its derivation, see [28, Proposition 2, Appendix A.2.1], [26, Section 3.4.3], [14, Section 2.5] and [12, Section 2.3]. Though the correction term in the TIC is robust under model misspecification, the term generally depends on unknown population parameters and tends to be relatively complicated. In practice, the correction term is given by its sample counterpart without changing the order of the remainder in approximation to the exact bias correction.

Improvements of the Akaike and Takeuchi information criteria have been given typically by deriving the higher-order asymptotic biases and the asymptotic biases of the estimated lower-order asymptotic biases. Konishi and Kitagawa [25] gave these results using the von Mises calculus [46,47] while Ogasawara [38, Section 3] obtained the general results using log likelihood derivatives and gave simplified formulas in the case of the exponential family of distributions under canonical parametrization.

E-mail address: hogasa@res.otaru-uc.ac.jp.

http://dx.doi.org/10.1016/j.jmva.2016.04.003
0047-259X/

A typical problem in model selection by the AIC is how to choose appropriate regressors among candidates in regression models. Sugiura [42] gave the exact correction term for the AIC in normal univariate linear regression under correct model specification. Fujikoshi and Satoh [16] derived the corresponding higher-order asymptotic bias under possible model misspecification in normal multivariate linear regression. Yanagihara, Sekiguchi and Fujikoshi [48] obtained the higher-order term in logistic regression. Kamo, Yanagihara and Satoh [22] derived the higher-order term in Poisson regression. Ogasawara [38, Section 6] added the corresponding results in gamma and negative binomial regression. It is known that in normal linear regression, the AIC tends to choose models with relatively large number of regressors or complicated models (see [18,16]). It has been shown that the bias-corrected AIC substantially increases the proportions of choosing correct models in normal linear regression [16].

The AIC has also been used in factor analysis [43] or more generally in covariance structure analysis and structural equation modeling. A typical problem in exploratory factor analysis (EFA) is to select an appropriate number of common factors. This problem partially motivated H. Akaike to coin the AIC [24, p. 124]. It is known as in regression analysis that in EFA the AIC tends to choose relatively excessive number of common factors [2, Table 5], [19, p. 39]. It is to be noted that the factor analysis model can be seen as a multivariate linear regression model, where regressor(s) are latent common factors. Note also that even in usual regression models error term(s) are latent variables corresponding to unique factors in EFA.

This paper gives the usual and higher-order correction terms of the TIC and AIC in EFA using the Wishart likelihood under possible non-normality. Simple approximations to the higher-order correction term in the AIC without using unknown population factor loadings and unique variances are also given under normality. The approximations are shown to be similar to the corresponding simulated and exact asymptotic values. In the simulation of model selection, it is shown that the proportion of choosing a correct number of common factors increases by the approximation to the higher-order correction term in the AIC under normality and different conditions of non-normality.

2. Bias correction of the AIC and TIC

Let

$$\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (i = 1, \dots, N) \quad \text{and} \quad \mathbf{U} \equiv \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})', \tag{2.1}$$

where $\mathbf{x}_i$ is the $p \times 1$ vector of observable variables, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are its expectation and covariance matrix, respectively. Then, it is known that $\mathbf{U}$ is Wishart distributed with the scale parameter $\boldsymbol{\Sigma}$ and N degrees of freedom, which is denoted by

$$\mathbf{U} \sim \mathrm{W}_p(\boldsymbol{\Sigma}, N), \tag{2.2}$$

whose probability density is

$$f(\mathbf{U}|\boldsymbol{\Sigma}) = \frac{|\mathbf{U}|^{(N-p-1)/2} \exp\{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{U})/2\}}{2^{pN/2} |\boldsymbol{\Sigma}|^{N/2} \Gamma_p(N/2)} \tag{2.3}$$

[3, Section 7.2], where $\Gamma_p(\cdot)$ is the p-variate gamma function given by

$$\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\{t - (1/2)(i-1)\} \tag{2.4}$$

and $\Gamma\{\cdot\}$ is the usual or univariate gamma function.

Let

$$\mathbf{S} = n^{-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \tag{2.5}$$

with $n = N - 1$, $\bar{\mathbf{x}} = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i$ and $\mathrm{E}_f(\mathbf{S}) = \boldsymbol{\Sigma}$, where $\mathrm{E}_f(\cdot)$ is an expectation under correct model (distribution) specification corresponding to (2.3) or equivalently to (2.1). Using the property of normally distributed variables with (2.3) and (2.5), it can be shown that

$$\mathbf{S} \sim \mathrm{W}_p(n^{-1}\boldsymbol{\Sigma}, n), \tag{2.6}$$

and its density is

$$f(\mathbf{S}|\boldsymbol{\Sigma}) = \frac{|\mathbf{S}|^{(n-p-1)/2} \exp\{-\mathrm{tr}(n\boldsymbol{\Sigma}^{-1}\mathbf{S})/2\}}{(2/n)^{pn/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma_p(n/2)}. \tag{2.7}$$