Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 133 (2005) 417–433

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On distribution of AIC in linear regression models

Hirokazu Yanagihara[a,*], Chihiro Ohmoto[b,1]

[a] *Department of Statistical Methodology, The Institute of Statistical Mathematics, Tokyo 106-8692, Japan*
[b] *Department of Mathematics, Faculty of School Education, Hiroshima University, Higashi-Hiroshima, Hiroshima 739-8524, Japan*

Received 13 September 2002; accepted 23 March 2004
Available online 15 June 2004

Abstract

This paper investigates an asymptotic distribution of the Akaike information criterion (AIC) and presents its characteristics in normal linear regression models. The bias correction of the AIC has been studied. It may be noted that the bias is only the mean, i.e., the first moment. Higher moments are important for investigating the behavior of the AIC. The variance increases as the number of explanatory variables increases. The skewness and kurtosis imply a favorable accuracy of the normal approximation. An asymptotic expansion of the distribution function of a standardized AIC is also derived.

MSC: primary 62E20; secondary 62B10

Keywords: Akaike information criterion; Asymptotic distribution; Asymptotic expansion; Asymptotic kurtosis; Asymptotic skewness; Asymptotic variance; Confidence interval; Linear regression model; Maximum likelihood estimator; Non-centrality parameter

1. Introduction

Consider the linear regression model given by

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

* Corresponding author. Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan. Tel.: +81-29-853-6451; fax: +81-29-853-6451.
E-mail address: yanagi@sk.tsukuba.ac.jp (H. Yanagihara)

[1] Current address: Yokota Junior High School, 2050-3 Inahara, Yokota-chou, Nita-gun, Shimane 699-1821, Japan.

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.03.016

where $\boldsymbol{y}=(y_1,\ldots,y_n)'$ is an observable random vector, n is the sample size, $X=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)'$ is an $n\times k$ matrix of explanatory variables with the full rank k, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_k)'$ is an unknown parameter vector, and $\boldsymbol{\epsilon}=(\varepsilon_1,\ldots,\varepsilon_n)'$ is an unobservable error vector which is distributed as $\mathrm{N}_n(\mathbf{0},\sigma^2 I_n)$.

There is the problem of selecting an appropriate explanatory variable, in other words, an appropriate model. The Akaike information criterion (AIC) proposed by Akaike (1973) is one of the typical model selection methods, which determine the best model to be the minimizer of an expected discrepancy. The AIC is an approximate unbiased estimator of the risk function and assumes that all of the candidate models contain the true model. Precise unbiased estimators were proposed by Sugiura (1978), Hurvich and Tsai (1989), and Fujikoshi and Satoh (1997), and so on. This research focuses only on the bias of the AIC, i.e., the mean of the AIC.

Higher moments of the AIC are important for investigating the behavior of the AIC. McQuarrie and Tsai (1998) calculated the variance of the difference between two AICs under the condition that two candidate models include the true model. Kishino and Hasegawa (1989) and Shimodaira (1993) proposed an estimator for the variance of the difference between two information criteria. Shimodaira (1993) further discussed the confidence interval by using its estimator. For a test to compare sizes with two AICs, Linhart (1988) and Shimodaira (1997) gave an asymptotic distribution of the standardized difference between two AICs. Also, Vuong (1989) calculated the variance of the likelihood ratio of the candidate model and true model.

The asymptotic variance can be obtained, including for the case where the model does not include the true model. The asymptotic variance increases as the number of explanatory variables increases. Furthermore, simulation studies show that, when the model does not include the true model, the increase is larger. The skewness and kurtosis imply a favorable accuracy of the normal approximation.

This paper is constructed as follows. In Section 2, the asymptotic variance of the AIC is obtained up to the order n^{-1}. In Section 3, the skewness and kurtosis of the standardized AIC (not two different AICs) are derived up to the order n^{-1}. Applications of our theoretical results to real data are described in Section 4.

2. Asymptotic variance of AIC

In this section, we obtain an expansion of the variance of AIC up to the second order term and discuss the characteristics of variance. We clarify the notations of the model used in this paper and consider the distribution of a maximum likelihood estimator (MLE) of σ^2 for a candidate model. By using this distribution and a perturbation expansion of AIC, we obtain an expansion of the variance of AIC for the candidate model up to o(1). Through the expansion, we consider the characteristics of the variance of AIC.

A generic candidate model can be expressed in terms of subset j of integers $\omega=\{1,\ldots,k\}$ and the matrix $X_{(j)}$ consisting of the columns of X indexed by the $k_{(j)}$ integers in j. A candidate model is expressed as

$$M_{(j)}:\boldsymbol{y}=X_{(j)}\boldsymbol{\beta}_{(j)}+\boldsymbol{\epsilon}_{(j)},$$