Journal of Multivariate Analysis 122 (2013) 239–250

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

High-dimensional AIC in the growth curve model

Yasunori Fujikoshi [a,*], Rie Enomoto [b], Tetsuro Sakurai [c]

[a] Department of Mathematics, Graduate School of Science, Hiroshima University, Higashi-Hiroshima, 739-8626, Japan
[b] Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601, Japan
[c] Center of General Education, Tokyo University of Science, Suwa, 5000-1 Toyohira, Chino, Nagano 391-0292, Japan

ARTICLE INFO

Article history:
Received 11 November 2012
Available online 27 July 2013

AMS 2000 subject classifications:
primary 62H12
secondary 62H30

Keywords:
AIC
HAIC
Asymptotic distributions
High-dimensional criteria
Growth curve model

ABSTRACT

The AIC and its modifications have been proposed for selecting the degree in a polynomial growth curve model under a large-sample framework when the sample size n is large, but the dimension p is fixed. In this paper, first we propose a high-dimensional AIC (denoted by HAIC) which is an asymptotic unbiased estimator of the AIC-type risk function defined by the expected log-predictive likelihood or equivalently the Kullback–Leibler information, under a high-dimensional framework such that $p/n \to c \in [0, 1)$. It is noted that our new criterion gives an estimator with small biases in a wide range of p and n. Next we derive asymptotic distributions of AIC and HAIC under the high-dimensional framework. Through a Monte Carlo simulation, we note that these new approximations are more accurate than the approximations based on a large-sample framework.

1. Introduction

We consider the growth curve model introduced by Potthoff and Roy [8], which is given by

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\Theta}\mathbf{X} + \mathcal{E}, \tag{1.1}$$

where $\mathbf{Y}$; $n \times p$ is an observation matrix, $\mathbf{A}$; $n \times q$ is a design matrix across individuals, $\mathbf{X}$; $k \times p$ is a design matrix within individuals, $\boldsymbol{\Theta}$ is an unknown matrix, and each row of $\mathcal{E}$ is independent and identically distributed as a p-dimensional normal distribution with mean $\mathbf{0}$ and an unknown covariance matrix $\boldsymbol{\Sigma}$. We assume that $n - p - q - 1 > 0$, and $\mathrm{rank}(\mathbf{X}) = k$. If we consider a polynomial regression of degree $k - 1$ on the time t and with q groups, then

$$\mathbf{A} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_q} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_p \\ \vdots & \vdots & \vdots & \vdots \\ t_1^{k-1} & t_2^{k-1} & \cdots & t_p^{k-1} \end{pmatrix}, \tag{1.2}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector whose elements are all one.

* Corresponding author.
E-mail address: fujikoshi_y@yahoo.co.jp (Y. Fujikoshi).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.07.006

It is important in a polynomial growth curve model to decide its degree. One way is to treat the problem as the one of selecting models. Related to such problems, consider a set of candidate models $M_1, \ldots, M_k$, where M_j is defined by

$$M_j;\ \mathbf{Y} = \mathbf{A}\boldsymbol{\Theta}_j\mathbf{X}_j + \mathcal{E}, \quad j = 1, \ldots, k. \tag{1.3}$$

Here $\boldsymbol{\Theta}_j$ is the $q \times j$ submatrix of $\boldsymbol{\Theta}$, and $\mathbf{X}_j$ is the $j \times p$ submatrix of $\mathbf{X}$ defined by

$$\boldsymbol{\Theta} = (\boldsymbol{\Theta}_j, \boldsymbol{\Theta}_{(j)}), \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_j \\ \mathbf{X}_{(j)} \end{pmatrix}.$$

The AIC [1] for M_j is given by

$$\mathrm{AIC} = n \log |\hat{\boldsymbol{\Sigma}}_j| + np(\log 2\pi + 1) + 2\left\{qj + \frac{1}{2}p(p+1)\right\}, \tag{1.4}$$

where $\hat{\boldsymbol{\Sigma}}_j$ is the MLE of $\boldsymbol{\Sigma}$ under M_j, which is given by

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n}(\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\Theta}}_j\mathbf{X}_j)'(\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\Theta}}_j\mathbf{X}_j),$$

where $\hat{\boldsymbol{\Theta}}_j = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{X}_j'(\mathbf{X}_j\mathbf{S}^{-1}\mathbf{X}_j')^{-1}$, $\mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{Y}/(n-q)$, and $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. The constant $\{qj + p(p+1)/2\}$ is the number of independent parameters under M_j. In addition to AIC, there are some modifications (see [10]) which were proposed as approximately unbiased estimators of the AIC-type risk function defined by the expected log-predictive likelihood or equivalently the Kullback–Leibler information, based on a large-sample theory. The modifications were studied assuming that the true model is included into the largest candidate model M_k.

In general, the approximations based on a large-sample framework become inaccurate as the dimension p increases while sample size n remains fixed. On the other hand, in last years we encounter more and more problems in applications when p is comparable with n or even exceeds it. So, it is important to examine behavior of AIC when the dimension is large, for example, a high-dimensional framework such that

$$n \to \infty, p \to \infty, \quad \frac{p}{n} \to c \in [0, 1). \tag{1.5}$$

In this paper we first derive a high-dimensional AIC denoted by HAIC which is an asymptotic unbiased estimator of the AIC-type risk under (1.5). It is noted that our new criterion gives a better estimator for the unbiasedness of the AIC-type risk in a wide range of p and n.

Recently, it has been noted [3,15] that the AIC and its modifications in multivariate regression model have a consistency property under a high-dimensional framework. However, we note that such property cannot be seen for AIC and HAIC of selecting the degree in the growth curve model.

Next, we derive asymptotic distributions of AIC and HAIC under (1.5). More precisely, let the values of AIC and HAIC for model M_j denote by AIC_j and HAIC_j, respectively, and then, the best subsets chosen by minimizing AIC and HAIC are written as

$$\hat{j}_\mathrm{A} = \arg\min_j \mathrm{AIC}_j, \qquad \hat{j}_\mathrm{HA} = \arg\min_j \mathrm{HAIC}_j.$$

Then we derive obtain asymptotic distributions of $\hat{j}_\mathrm{A}$ and $\hat{j}_\mathrm{HA}$ under (1.5). The results include the large-sample asymptotic distributions as their special cases. Through a simulation experiment, the high-dimensional asymptotic approximations are more accurate than the large-sample approximations in a wide range of p and n. We also point some tendencies of selecting the true model by AIC and HAIC, based on a simulation experiment.

In this paper we consider the case when $\mathbf{A}$ and $\mathbf{X}$ are given in (1.2) under the growth curve model. However, our results can be easily extended to the case that $\mathbf{A}$ and $\mathbf{X}$ are general.

The present paper is organized as follows: in Section 2, we present the AIC-type risk and the bias term. Further, we give a distributional reduction for the bias term. In Section 3, we propose HAIC. In Section 4, we derive asymptotic distributions of AIC and HAIC under a high-dimensional framework. In Section 5, we verify that HAIC has smaller bias than AIC as the estimators of AIC-type risk, and the high-dimensional approximations are more accurate than the large-sample approximations, by conducting numerical experiments.

2. Preliminaries

As is well known, the AIC was proposed as an approximately unbiased estimator of the AIC-type risk function defined by the expected log-predictive likelihood or equivalently the Kullback–Leibler information. Let $f(\mathbf{Y}; \boldsymbol{\Theta}_j, \boldsymbol{\Sigma}_j)$ be the density function of $\mathbf{Y}$ under M_j. Then the expected log-predictive likelihood of M_j is defined by

$$R_A = \mathrm{E}^*_\mathbf{Y}\mathrm{E}^*_{\mathbf{Y}_F}[-2\log f(\mathbf{Y}_F; \hat{\boldsymbol{\Theta}}_j, \hat{\boldsymbol{\Sigma}}_j)], \tag{2.1}$$