GENERAL FORMULA OF BIAS-CORRECTED AIC IN GENERALIZED LINEAR MODELS

Shinpei Imori[1], Hirokazu Yanagihara, Hirofumi Wakaki

Department of Mathematics, Graduate School of Science, Hiroshima University

ABSTRACT

The present paper considers a bias correction of Akaike's information criterion (AIC) for selecting variables in the generalized linear model (GLM). When the sample size is not so large, the AIC has a non-negligible bias that will negatively affect variable selection. In the present study, we obtain a simple expression for a bias-corrected AIC (corrected AIC, or CAIC) in GLMs. A numerical study reveals that the CAIC has better performance than the AIC for variable selection.

Key words: Akaike's information criterion, Bias correction, Generalized linear model, Maximum likelihood estimation, Variable selection.

1. INTRODUCTION

In real data analysis, deciding the best subset of variables in regression models is an important problem. It is common for a model selection to measure the goodness of fit of the model by the risk function based on the expected Kullback-Leibler (KL) information (Kullback and Leibler (1951)). For actual use, we must estimate the risk function, which depends on unknown parameters. The most famous estimator of the risk function is Akaike's information criterion (AIC) proposed by Akaike (1973). Since the AIC can be simply defined as $-2 \times$ the maximum log-likelihood $+2 \times$ the number of parameters, the AIC is widely applied in chemometrics, engineering, econometrics, psychometrics, and many other fields for selecting appropriate models using a set of explanatory variables.

In addition, the order of the bias of the AIC to the risk function is $O(n^{-1})$, which indicates implicitly that the AIC sometimes has a non-negligible bias to the risk function when the sample size n is not so large. The AIC tends to underestimate the risk function and the bias of AIC tends to increase with the number of parameters in the model. Potentially, the AIC tends to choose the model that has more parameters than the true model as the best model (see Shibata (1980)). Combined with these characteristics, the bias will cause a disadvantage whereby the model having the most parameters is easily chosen by the AIC among the candidate models as the best model. One method of avoiding this undesirable result is to correct the bias of the AIC. A number of authors

[1]Corresponding author
E-mail address: m106547@hiroshima-u.ac.jp

have investigated bias-corrected AIC for various models. For example, Sugiura (1978) developed an unbiased estimator of the risk function in linear regression models, which is the UMVUE of the risk function reported by Davies *et al.* (2006). Hurvich and Tsai (1989) formally adjusted the bias of the AIC (called AICc) in several models. In particular, the AICc is equivalent to Sugiura's bias-corrected AIC in the case of the linear regression model. Wong and Li (1998) extended Hurvich and Tsai's AICc to a wider model and verified that their AICc has a higher performance than the original AIC by conducting numerical studies.

Unfortunately, except for the linear regression model, the AICc does not completely reduce the bias of the AIC to $O(n^{-2})$. As mentioned previously, the goodness of fit of the model is measured by the risk function based on the expected KL information. Thus, obtaining a higher-order asymptotic unbiased estimator of the risk function will allow us to correctly measure the goodness of fit of the model. This will further facilitate the reasonable selection of variables. From this viewpoint, Yanagihara *et al*. (2003) and Kamo *et al*. (2011) proposed the bias-corrected AIC's in the logistic model and the Poisson regression model, respectively, which complementarily reduce the bias of the AIC to $O(n^{-2})$. We refer to the completely bias-corrected AIC to $O(n^{-2})$ as the corrected AIC (CAIC). Frequently, the CAIC improves the performance of the original AIC dramatically. This strongly suggests the usefulness of the CAIC for real data analysis. Nevertheless, the CAIC is rarely used in real data analysis because the CAIC has been derived only in a few models. Moreover, since the derivation of the CAIC is complicated, a great deal of practice is needed in order to carry out the calculation of the CAIC if a researcher wants to use the CAIC in a model in which the CAIC has not been derived. Hence, the CAIC is not a user-friendly model selector under the present circumstances, although the CAIC has better performance than the original AIC. If we can obtain the CAIC in a small amount of time, the CAIC will become a useful and user-friendly model selector.

The goal of the present paper is to derive a simple formula for the CAIC in the widest model possible. The model considered is the generalized linear model (GLM), which was proposed by Nelder and Wedderburn (1972). Nevertheless, the GLM can express a number of statistical models by changing the distribution and the link function, such as the normal linear regression model, the logistic regression model, and the probit model, which are currently commonly used in a number of applied fields, cf., Barnett and Nurmagambetov (2010), Matas *et al*. (2010), Sánchez-Carneo *et al*. (2011), and Teste and Lieffers (2011). Practically speaking, the GLM can be easily fitted to real data using the "glm" function in "R" (R Development Core Team (2011)), which implements a number of distributions and link functions. Since the model considered herein is wide and can be easily fitted to real data, the CAIC in the GLM is confirmed useful in real data analysis. Generally, the additional bias correction terms in the CAIC requires estimators of several orders of moments of the response variables. Such moments should be calculated for each specified model. However, we emphasize that the moments do not remain in our formula. Hence, in order to apply the CAIC to the real data with our formula, we need only calculate the derivatives of the log-likelihood function of less than or equal to the fourth order. Our formula can be applied using formula manipulation software such as “Mathematica”.

The remainder of the present paper is organized as follows: In Section 2, we consider a stochastic expansion of the maximum likelihood estimator (MLE) in the GLM. In Section 3, we propose a new information criterion by complementarily reducing the bias of the AIC in the GLMs to $O(n^{-2})$. In Section 4, we present several examples of the CAIC in a model in which the “glm” program is implemented. These examples will be helpful to applied researchers. In Section 5, we investigate the performance of the proposed CAIC through numerical simulations. Technical details are provided in the Appendix.

2. STOCHASTIC EXPANSION OF THE MLE IN THE GLM

The GLM considered herein is developed to allow us to fit regression models for the response variables that follow a very general distribution belonging to the exponential family, the probability density function of which is given as follows:

$$f(y;\theta,\phi)=\exp\left\{\frac{\theta y-b(\theta)}{a(\phi)}+c(y,\phi)\right\}, \tag{2.1}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are known functions, the unknown parameter θ is referred to as the natural location parameter, and ϕ is often referred to as the scale parameter. (For the details of the GLM, see, e.g., Meyer *et al*. (2002).) In the present paper, we assume that ϕ is known. The exponential family includes the normal, binomial, Poisson, geometric, negative binomial, exponential, gamma, and inverse normal distributions. Let each datum consist of a sequence $\{(y_i,\boldsymbol{x}_i);(i=1,\ldots,n)\}$, where $y_1,\ldots,y_n$ are independent random variables referred to as response variables, and $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are p-dimensional nonstochastic vectors referred to as explanatory variables. The expectation of the response y_i is related to a linear predictor $\eta_i=\boldsymbol{x}_i'\boldsymbol{\beta}$ by a link function $h(\cdot)$, i.e., $h(\mathrm{E}[y_i])=h(\mu(\theta_i))=\eta_i$. For theoretical purposes, we define $u=(h\circ\mu)^{-1}$, i.e., $\theta_i=u(\eta_i)$. When $h=\mu^{-1}$, i.e., u is an identity function, we say that h is the natural link function. For example, the logistic regression model uses the natural link function. Finally, the candidate model is expressed as

$$y_i\overset{\text{i.d}}{\sim}f(y_i;\theta_i(\boldsymbol{\beta}),\phi),$$

where $f(\cdot)$ is given by (2.1). The p-dimensional unknown vector $\boldsymbol{\beta}$ can be estimated by the maximum likelihood method. The joint probability density function of $\boldsymbol{y}=(y_1,\ldots,y_n)'$ is given by

$$f(\boldsymbol{y};\boldsymbol{\beta})=\prod_{i=1}^{n}f(y_i;\theta_i(\boldsymbol{\beta}),\phi)=\prod_{i=1}^{n}\exp\left\{\frac{\theta_i y_i-b(\theta_i)}{a(\phi)}+c(y_i,\phi)\right\}.$$

Hence, the log-likelihood function of the GLM is expressed as

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) = \log f(\boldsymbol{y}; \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}. \tag{2.2}$$

Let $\hat{\boldsymbol{\beta}}$ be the MLE of $\boldsymbol{\beta}$. Here, $\hat{\boldsymbol{\beta}}$ is given as the solution of the following likelihood equation:

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\beta}; \boldsymbol{y}) = \frac{1}{a(\phi)} \sum_{i=1}^{n} \left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i} \right) \frac{\partial \theta_i}{\partial \eta_i} \boldsymbol{x}_i = \frac{1}{a(\phi)} X' \Delta (\boldsymbol{y} - \boldsymbol{\mu}) = \mathbf{0}_p,$$

where $X = (\boldsymbol{x}_1, \ldots \boldsymbol{x}_n)'$, $\Delta = \mathrm{diag}(\partial\theta_1/\partial\eta_1, \ldots, \partial\theta_n/\partial\eta_n)$, and $\boldsymbol{\mu} = (\partial b(\theta_1)/\partial\theta_1, \ldots, \partial b(\theta_n)/\partial\theta_n)'$. Note that $b(\theta)$ is a C^∞-class function and all of the orders of the moments of $\boldsymbol{y}$ exist in the interior Θ^0 of the natural parameter space Θ. In using some of the properties of the MLE, we have the following regularity assumptions (see, e.g., Fahrmeir and Kaufmann (1985)):

C.1 : $\boldsymbol{x}_i'\boldsymbol{\beta} \in h(\mathcal{M})$ $(i = 1, \ldots, n)$, for all $\boldsymbol{\beta} \in \mathcal{B}$,

C.2 : h is three times continuously differentiable,

C.3 : For all $\boldsymbol{x}_i \in \chi$, $\partial\theta_i/\partial\eta_i \neq 0$, $(i = 1, \ldots, n)$,

C.4 : $^\exists n_0$ *s.t.* $X'X$ has full rank for $n \geq n_0$,

where $\mathcal{B}$ is an admissible open set in $\mathbb{R}^p$ for the parameter $\boldsymbol{\beta}$, χ is a compact set for the regressors $\boldsymbol{x}_i$, and $\mathcal{M}$ denotes the image $\mu(\Theta^0)$. Condition C.1 is necessary in order to obtain the GLM for all $\boldsymbol{\beta}$. Condition C.2 is necessary in order to calculate the bias. Conditions C.3 and 4 ensure that $X'\Delta V \Delta X$ is positive definite for all $\boldsymbol{\beta} \in \mathcal{B}$, $n \geq n_0$, where

$$V = a(\phi)\mathrm{diag}\left(\frac{\partial^2 b(\theta_1)}{\partial \theta_1^2}, \ldots, \frac{\partial^2 b(\theta_n)}{\partial \theta_n^2} \right).$$

Moreover, we have the following additional conditions to assure strong consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$, which can be derived by slightly modifying the results reported by Fahrmeir and Kaufmann (1985):

C.5 : sequence $\{\boldsymbol{x}_i\}$ lies in χ with $u(\boldsymbol{x}'\boldsymbol{\beta}) \in \Theta^0$, $\boldsymbol{\beta} \in \mathcal{B}$,

C.6 : $\liminf_{n\to\infty} \lambda_{\min}(X'\Delta V \Delta X/n) > 0$,

C.7 : $^\exists c > 0$, n_1, $\lambda_{\min}(X'X) > c\lambda_{\max}(X'X)$, $n \geq n_1$,

where $\lambda_{\min}(A)$ is the smallest eigenvalue of symmetric matrix A. According to Theorem 5 in Fahrmeir and Kaufmann (1985), $\hat{\boldsymbol{\beta}}$ has strong consistency and asymptotic normality under these conditions. Furthermore, from C.6, $X'\Delta V \Delta X = O(n)$, $(n \to \infty)$.

Based on the above conditions, $\hat{\boldsymbol{\beta}}$ can be formally expanded as follows:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \frac{1}{\sqrt{n}}\boldsymbol{b}_1 + \frac{1}{n}\boldsymbol{b}_2 + \frac{1}{n\sqrt{n}}\boldsymbol{b}_3 + O_p(n^{-2}). \tag{2.3}$$

Note that $\partial\ell(\hat{\boldsymbol{\beta}};\boldsymbol{y})/\partial\boldsymbol{\beta} = \boldsymbol{0}_p$. By applying a Taylor expansion around $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ to this equation, the likelihood equation is expanded as follows:

$$\begin{aligned}\boldsymbol{0}_p &= \frac{1}{\sqrt{n}}(\boldsymbol{g} + G_2\boldsymbol{b}_1) + \frac{1}{n}\left\{G_1\boldsymbol{b}_2 + \frac{1}{2}G_3(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)\right\}\\ &\quad + \frac{1}{n\sqrt{n}}\left\{G_2\boldsymbol{b}_3 + \frac{1}{2}G_3(\boldsymbol{b}_1 \otimes \boldsymbol{b}_2 + \boldsymbol{b}_2 \otimes \boldsymbol{b}_1) + \frac{1}{6}(I_p \otimes \boldsymbol{b}_1')G_4(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)\right\} + O_p(n^{-2}),\end{aligned} \tag{2.4}$$

where $\boldsymbol{0}_p$ is a p-dimensional vector of zeros, and

$$\begin{aligned}\boldsymbol{g} &= \frac{1}{\sqrt{n}}\frac{\partial\ell(\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\beta}} = \frac{1}{\sqrt{n}a(\phi)}\sum_{i=1}^{n}(y_i - d_{i1})c_{i1}\boldsymbol{x}_i,\\ G_2 &= \frac{1}{n}\frac{\partial^2\ell(\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \frac{1}{na(\phi)}\sum_{i=1}^{n}\{-d_{i2}c_{i1}^2 + (y_i - d_{i1})c_{i2}\}\boldsymbol{x}_i\boldsymbol{x}_i',\\ G_3 &= \frac{1}{n}\left(\frac{\partial}{\partial\boldsymbol{\beta}'} \otimes \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)\ell(\boldsymbol{\beta};\boldsymbol{y}),\\ &= \frac{1}{na(\phi)}\sum_{i=1}^{n}\{-d_{i3}c_{i1}^3 - 3d_{i2}c_{i1}c_{i2} + (y_i - d_{i1})c_{i3}\}(\boldsymbol{x}_i' \otimes \boldsymbol{x}_i\boldsymbol{x}_i'),\\ G_4 &= \frac{1}{n}\left(\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \otimes \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)\ell(\boldsymbol{\beta};\boldsymbol{y})\\ &= \frac{1}{na(\phi)}\sum_{i=1}^{n}\{-d_{i4}c_{i1}^4 - 6d_{i3}c_{i1}^2c_{i2} - 3d_{i2}c_{i1}c_{i2}^2 - 4d_{i2}c_{i3} + (y_i - d_{i1})c_{i4}\}(\boldsymbol{x}_i\boldsymbol{x}_i' \otimes \boldsymbol{x}_i\boldsymbol{x}_i').\end{aligned}$$

Here, coefficients c_{ik} and d_{ik}, $(i = 1,\ldots,n; k = 1,\ldots,4)$ are defined as

$$c_{ik} = \frac{\partial^k\theta_i}{\partial\eta_i^k}, \quad d_{ik} = \frac{\partial^k b(\theta_i)}{\partial\theta_i^k}. \tag{2.5}$$

Note that c_{ik} is determined by the link function, and d_{ik} is determined by the distribution of the model. Let us define $Z_i = \sqrt{n}(G_i - M_i)$ $(i = 2, 3, 4)$, where $M_i = \mathrm{E}[G_i]$, the explicit forms of which are

$$\begin{aligned}M_2 &= \frac{1}{na(\phi)}\sum_{i=1}^{n}(-d_{i2}c_{i1}^2)\boldsymbol{x}_i\boldsymbol{x}_i',\\ M_3 &= \frac{1}{na(\phi)}\sum_{i=1}^{n}(-d_{i3}c_{i1}^3 - 3d_{i2}c_{i1}c_{i2})(\boldsymbol{x}_i' \otimes \boldsymbol{x}_i\boldsymbol{x}_i'),\\ M_4 &= \frac{1}{na(\phi)}\sum_{i=1}^{n}(-d_{i4}c_{i1}^4 - 6d_{i3}c_{i1}^2c_{i2} - 3d_{i2}c_{i2}^2 - 4d_{i2}c_{i1}c_{i3})(\boldsymbol{x}_i\boldsymbol{x}_i' \otimes \boldsymbol{x}_i\boldsymbol{x}_i').\end{aligned}$$

Thus, Z_i $(i = 2, 3, 4)$ can be expressed as

$$Z_2 = \frac{1}{\sqrt{n}a(\phi)} \sum_{i=1}^{n} \{(y_i - d_{i1})c_{i2}\} \boldsymbol{x}_i \boldsymbol{x}_i',$$
$$Z_3 = \frac{1}{\sqrt{n}a(\phi)} \sum_{i=1}^{n} \{(y_i - d_{i1})c_{i3}\} (\boldsymbol{x}_i' \otimes \boldsymbol{x}_i \boldsymbol{x}_i'),$$
$$Z_4 = \frac{1}{\sqrt{n}a(\phi)} \sum_{i=1}^{n} \{(y_i - d_{i1})c_{i4}\} (\boldsymbol{x}_i \boldsymbol{x}_i' \otimes \boldsymbol{x}_i \boldsymbol{x}_i').$$

Based on the regularity assumptions, nonsingularity of M_2 is guaranteed. Furthermore, the regularity assumptions and Conditions C.6 and C.7 ensure the asymptotic normality of Z_i. Hence, we can rewrite (2.4) as

$$\begin{aligned} \boldsymbol{0}_p = & \frac{1}{\sqrt{n}} (\boldsymbol{g} + M_2 \boldsymbol{b}_1) + \frac{1}{n} \left\{ M_2 \boldsymbol{b}_2 + \frac{1}{2} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) + Z_2 \boldsymbol{b}_1 \right\} \\ & + \frac{1}{n\sqrt{n}} \left\{ M_2 \boldsymbol{b}_3 + \frac{1}{2} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_2 + \boldsymbol{b}_2 \otimes \boldsymbol{b}_1) \right. \\ & \left. + \frac{1}{6} (I_p \otimes \boldsymbol{b}_1') M_4 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) + Z_2 \boldsymbol{b}_2 + \frac{1}{2} Z_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) \right\} + O_p(n^{-2}). \end{aligned} \quad (2.6)$$

Comparing the terms of the same order in both sides of (2.6), the explicit forms of $\boldsymbol{b}_1, \boldsymbol{b}_2$, and $\boldsymbol{b}_3$ are obtained as follows:

$$\begin{aligned} \boldsymbol{b}_1 &= -M_2^{-1} \boldsymbol{g}, \\ \boldsymbol{b}_2 &= -M_2^{-1} \left\{ \frac{1}{2} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) + Z_2 \boldsymbol{b}_1 \right\}, \\ \boldsymbol{b}_3 &= -M_2^{-1} \left\{ \frac{1}{2} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_2 + \boldsymbol{b}_2 \otimes \boldsymbol{b}_1) + \frac{1}{6} (I_p \otimes \boldsymbol{b}_1)' M_4 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) + Z_2 \boldsymbol{b}_2 + \frac{1}{2} Z_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) \right\}. \end{aligned}$$

3. BIAS CORRECTION OF THE AIC

The goodness of fit of the model is measured by the risk function based on the expected KL information, as follows:

$$R = \mathrm{E}_{\boldsymbol{y}} \mathrm{E}_{\boldsymbol{y}^*} [-2\ell(\hat{\boldsymbol{\beta}}; \boldsymbol{y}^*)],$$

where $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)'$ is an n-dimensional random vector that is independent of $\boldsymbol{y}$ and has the same distribution as $\boldsymbol{y}$. At the beginning of this section, we derive the bias of $-2\ell(\hat{\boldsymbol{\beta}}; \boldsymbol{y})$ to R. Under ordinary circumstances, calculation of the expectations of $\boldsymbol{y}$ under the specific distribution are needed in order to express the bias. However, based on the characteristics of the exponential family, we can obtain the bias without calculating the expectations of $\boldsymbol{y}$ under the specific distribution. The explicit form of the bias can be expressed by several derivatives of the log-likelihood function.

The bias when we estimate R by $-2\ell(\hat{\boldsymbol{\beta}};\boldsymbol{y})$ is given as

$$B = R - \mathrm{E}_{\boldsymbol{y}}[-2\ell(\hat{\boldsymbol{\beta}};\boldsymbol{y})] = \mathrm{E}_{\boldsymbol{y}}\left[\mathrm{E}_{\boldsymbol{y}^*}\left[2\ell(\hat{\boldsymbol{\beta}};\boldsymbol{y}) - 2\ell(\hat{\boldsymbol{\beta}};\boldsymbol{y}^*)\right]\right] = 2\mathrm{E}_{\boldsymbol{y}}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}\hat{\theta}_i\right]. \tag{3.1}$$

By applying a Taylor expansion around $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ to $\hat{\theta}_i = (h\circ\mu)^{-1}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$, $\hat{\theta}_i$ is expanded as

$$\begin{aligned}\hat{\theta}_i = \theta_i &+ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\frac{\partial\theta_i}{\partial\boldsymbol{\beta}} + \frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\frac{\partial^2\theta_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &+ \frac{1}{6}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\left\{\left(\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)\theta_i\right\}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\otimes(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\} + O_p(n^{-2}). \end{aligned}\tag{3.2}$$

Substituting the stochastic expansion of $\hat{\boldsymbol{\beta}}$ in (2.3) into (3.2) yields the following:

$$\begin{aligned}\hat{\theta}_i = \theta_i &+ \frac{1}{\sqrt{n}}c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_1 + \frac{1}{n}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_2 + \frac{1}{2}c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)^2\right\} \\ &+ \frac{1}{n\sqrt{n}}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_3 + c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)(\boldsymbol{x}_i'\boldsymbol{b}_2) + \frac{1}{6}c_{i3}(\boldsymbol{x}_i'\boldsymbol{b}_1)^3\right\} + O_p(n^{-2}). \end{aligned}\tag{3.3}$$

By combining (3.1) and (3.3), we obtain

$$\begin{aligned}B = 2\mathrm{E}&\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}\theta_i\right] + \frac{2}{\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_1\right] \\ &+ \frac{2}{n}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_2 + \frac{1}{2}c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)^2\right\}\right] \\ &+ \frac{2}{n\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_3 + c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)(\boldsymbol{x}_i'\boldsymbol{b}_2) + \frac{1}{6}c_{i3}(\boldsymbol{x}_i'\boldsymbol{b}_1)^3\right\}\right] + O(n^{-2}). \end{aligned}\tag{3.4}$$

Recall that $d_{i1} = \partial b(\theta_i)/\partial\theta_i = \mathrm{E}[y_i]$. This yields the first term of (3.4), as follows:

$$2\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}\theta_i\right] = 0. \tag{3.5}$$

Since $\mathrm{E}[\boldsymbol{g}\boldsymbol{g}'] = -M_2$, the second term of (3.4) can be calculated as

$$\frac{2}{\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i - d_{i1})}{a(\phi)}c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_1\right] = -2\mathrm{E}[\boldsymbol{g}'M_2^{-1}\boldsymbol{g}] = 2p. \tag{3.6}$$

The third term of (3.4) can be obtained as

$$\begin{aligned}\mathrm{E}&\left[\frac{2}{n}\sum_{i=1}^{n}\frac{y_i - d_{i1}}{a(\phi)}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_2 + \frac{1}{2}c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)^2\right\}\right] \\ &= \frac{3}{n^2a(\phi)}\sum_{i=1}^{n}d_{i3}c_{i1}^2c_{i2}U_{ii}^2 + \frac{1}{n^3a(\phi)^2}\sum_{i,j}^{n}d_{i3}c_{i1}^3(d_{j3}c_{j1}^3 + 3d_{j2}c_{j1}c_{j2})U_{ij}^3 + O(n^{-2}), \end{aligned}\tag{3.7}$$

where $\sum_{i,j}^{n}$ refers to $\sum_{i=1}^{n}\sum_{j=1}^{n}$, and U_{ij} is the (i,j)th element of the matrix $U = XM_2^{-1}X'$, i.e.,

$$U_{ij} = \boldsymbol{x}_i'M_2^{-1}\boldsymbol{x}_j. \tag{3.8}$$

Note that coefficient U_{ij} is determined by both the link function and the distribution of the model. The derivation of (3.7) is shown in Appendix A.1. Furthermore, the fourth term of (3.4) can be expanded as

$$
\begin{aligned}
&\frac{2}{n\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i-d_{i1})}{a(\phi)}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_3+c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)(\boldsymbol{x}_i'\boldsymbol{b}_2)+\frac{1}{6}c_{i3}(\boldsymbol{x}_i'\boldsymbol{b}_1)^3\right\}\right]\\
&=-\frac{1}{n^2a(\phi)}\sum_{i=1}^{n}(d_{i4}c_{i1}^4+6d_{i3}c_{i1}^2c_{i2}-d_{i2}c_{i2}^2)U_{ii}^2\\
&\quad-\frac{1}{n^3a(\phi)^2}\sum_{i,j}^{n}\{2(d_{i3}c_{i1}^3)(d_{j3}c_{j1}^3+3d_{j2}c_{j1}c_{j2})+4(d_{i2}c_{i1}c_{i2})(d_{j2}c_{j1}c_{j2})\}U_{ij}^3\\
&\quad-\frac{1}{n^3a(\phi)^2}\sum_{i,j}^{n}\{(d_{i3}c_{i1}^3)(d_{j3}c_{j1}^3+3d_{j2}c_{j1}c_{j2})+4(d_{i2}c_{i1}c_{i2})(d_{j2}c_{j1}c_{j2})\}U_{ij}U_{ii}U_{jj}+O(n^{-2}). \qquad (3.9)
\end{aligned}
$$

The detailed derivation of (3.9) is given in Appendix A.2.

Finally, by substituting (3.5), (3.6), (3.7), and (3.9) into (3.4), we obtain the asymptotic expansion of B up to order n^{-1} as

$$B=2p+\frac{1}{n}(w_1+w_2)+O(n^{-2}), \qquad (3.10)$$

where

$$
\begin{aligned}
w_1&=-\frac{1}{na(\phi)}\sum_{i=1}^{n}(d_{i4}c_{i1}^4+3d_{i3}c_{i1}^2c_{i2}-d_{i2}c_{i2}^2)U_{ii}^2,\\
w_2&=-\frac{1}{n^2a(\phi)^2}\sum_{i,j}^{n}\{d_{i3}c_{i1}^3(d_{j3}c_{j1}^3+3d_{j2}c_{j1}c_{j2})+4(d_{i2}c_{i1}c_{i2})(d_{j2}c_{j1}c_{j2})\}(U_{ij}^3+U_{ij}U_{ii}U_{jj}).
\end{aligned} \qquad (3.11)
$$

By a simple calculation, we have $c_{i1}=1$ and $c_{i2}=0$ when the link function is natural. Thus, if the model has the natural link function, w_1 and w_2 became simple, as follows:

$$
\begin{aligned}
w_1&=-\frac{1}{na(\phi)}\sum_{i=1}^{n}d_{i4}U_{ii}^2,\\
w_2&=-\frac{1}{n^2a(\phi)^2}\sum_{i,j}^{n}d_{i3}d_{j3}(U_{ij}^3+U_{ij}U_{ii}U_{jj}).
\end{aligned} \qquad (3.12)
$$

Equation (3.10) yields the following formula for the CAIC:

$$\mathrm{CAIC}=\mathrm{AIC}+\frac{1}{n}(\hat{w}_1+\hat{w}_2), \qquad (3.13)$$

where $\hat{w}_1$ and $\hat{w}_2$ are defined by replacing $\boldsymbol{\beta}$ in w_1 and w_2 with $\hat{\boldsymbol{\beta}}$. On the other hand, if h is not the natural link function, we have to use w_1 and w_2 in (3.11). Note that $\hat{w}_1$ and $\hat{w}_2$ depend only on several derivatives. Therefore, we can comfortably obtain coefficients $\hat{w}_1$ and $\hat{w}_2$ using formula manipulation software.

4. THE CAIC IN THE MODELS IMPLEMENTED IN "glm"

In this section, we will present several examples of the CAIC in the GLM, which can be used in the "glm" of the "R" software. In "glm", the binomial distribution accepts the links "logit", "Probit", "cauchit", and "cloglog" (complementary log-log). The Gamma distribution accepts the "inverse", "identity", and "log" links. The Poisson accepts the distribution of the "log", "identity", and "sqrt" links, and the inverse Gaussian distribution accepts the "$1/\mu^2$", "inverse", and "log" links. These examples are obtained through our formula in (3.13). If h is the natural link function, w_1 and w_2 in (3.13) are expressed as shown in (3.12). Otherwise, w_1 and w_2 in (3.13) are expressed as in (3.11). Next, we present the coefficients of the CAIC in the all of models mentioned above. The results indicate that the CAIC in the model with a non-natural link function is more complex than that in the model with the natural link function. An expression of the log-likelihood function of the GLM is given by Equation (2.1).

4.1. Case of the Binomial Distribution

When we assume that y_i is distributed according to the Binomial distribution $B(m_i, p_i)$ $(i = 1, \ldots, n)$, the parameters and functions based on the distribution are given by

$$\theta_i = \log\left(\frac{p_i}{1-p_i}\right), \quad b(\theta_i) = m_i \log(1+\exp(\theta_i)),$$
$$\phi = 1, \quad a(\phi) = 1, \quad c(y_i, \phi) = \log\binom{m_i}{y_i}.$$

Then, the coefficients of the CAIC specifying the distribution are given by

$$d_{i1} = \frac{m_i \exp(\theta_i)}{1+\exp(\theta_i)}, \quad d_{i2} = \frac{m_i \exp(\theta_i)}{(1+\exp(\theta_i))^2},$$
$$d_{i3} = \frac{m_i \exp(\theta_i)(1-\exp(\theta_i))}{(1+\exp(\theta_i))^3}, \quad d_{i4} = \frac{m_i \exp(\theta_i)(1-4\exp(\theta_i)+\exp(2\theta_i))}{(1+\exp(\theta_i))^4}.$$

The remaining coefficients of the CAIC, which are determined by the link function, are as follows:

- Case of the logistic link function, i.e., $\mathrm{E}[y_i] = m_i p_i = m_i(1+\exp(-\eta_i))^{-1}$ (natural link function):

$$U_{ij} = \boldsymbol{x}_i' \left\{ -\frac{1}{n} \sum_{k=1}^{n} \frac{m_k \exp(-\eta_k)}{(1+\exp(-\eta_k))^2} \boldsymbol{x}_k \boldsymbol{x}_k' \right\} \boldsymbol{x}_j.$$

  The CAIC derived from the above coefficients coincides with the CAIC reported by Yanagihara *et al*. (2003).

- Case of the probit link function, i.e., $m_i p_i = m_i \Phi(\eta_i)$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution:

$$c_{i1} = \frac{\phi(\eta_i)}{(1-\Phi(\eta_i))\Phi(\eta_i)}, \quad c_{i2} = -\frac{\phi(\eta_i)}{\{(1-\Phi(\eta_i))\Phi(\eta_i)\}^2}\{\eta_i\Phi(\eta_i)(1-\Phi(\eta_i)) + \phi(\eta_i)(1-2\Phi(\eta_i))\},$$
$$U_{ij} = \boldsymbol{x}_i'\left\{-\frac{1}{n}\sum_{k=1}^{n}\frac{m_k\phi(\eta_k)^2}{(1-\Phi(\eta_k))\Phi(\eta_k)}\boldsymbol{x}_k\boldsymbol{x}_k'\right\}^{-1}\boldsymbol{x}_j,$$

where $\phi(\cdot)$ is the probability density function (PDF) of the standard normal distribution.

- Case of the cauchit link function, i.e., $m_ip_i = m_i\Psi(\eta_i)$, where $\Psi(\cdot)$ is the CDF of the standard Cauchy distribution:

$$c_{i1} = \frac{\psi(\eta_i)}{(1-\Psi(\eta_i))\Psi(\eta_i)}, \quad c_{i2} = \frac{\psi(\eta_i)^2}{\{(1-\Psi(\eta_i))\Psi(\eta_i)\}^2}\{2\pi\eta_i\Psi(\eta_i)(1-\Psi(\eta_i)) - (1-2\Psi(\eta_i))\},$$
$$U_{ij} = \boldsymbol{x}_i'\left\{-\frac{1}{n}\sum_{k=1}^{n}\frac{m_k\psi(\eta_k)^2}{(1-\Psi(\eta_k))\Psi(\eta_k)}\boldsymbol{x}_k\boldsymbol{x}_k'\right\}^{-1}\boldsymbol{x}_j,$$

where $\psi(\cdot)$ is the PDF of the standard Cauchy distribution.

- Case of the cloglog link function i.e., $m_ip_i = m_i - m_i\exp\{-\exp(\eta_i)\}$:

$$c_{i1} = \frac{\exp(\eta_i)}{1-\exp(-\exp(\eta_i))}, \quad c_{i2} = \frac{\exp(\eta_i)\{1-(1-\exp(\eta_i))\exp(-\exp(\eta_i))\}}{\{1-\exp(-\exp(\eta_i))\}^2},$$
$$U_{ij} = \boldsymbol{x}_i'\left\{-\frac{1}{n}\sum_{k=1}^{n}\frac{m_k\exp(2\eta_k)\exp(\exp(\eta_k))}{1-\exp(-\exp(\eta_k))}\boldsymbol{x}_k\boldsymbol{x}_k'\right\}^{-1}\boldsymbol{x}_j.$$

4.2. Case of the Poisson Distribution

Second, when we assume that y_i is distributed according to a Poisson distribution $Po(\lambda_i)$ ($i = 1, \ldots, n$), the parameters and functions based on the model are given as follows:

$$\theta_i = \log\lambda_i, \quad b(\theta_i) = \exp(\theta_i),$$
$$\phi = 1, \quad a(\phi) = 1, \quad c(y_i, \phi) = -\log(y_i!).$$

The coefficients of the CAIC specifying the distribution are given by

$$d_{i1} = d_{i2} = d_{i3} = d_{i4} = \exp(\theta_i).$$

The remaining coefficients of the CAIC, which are determined by the link function, are as follows:

- Case of the log link function, i.e., $\mathrm{E}[y_i] = \lambda_i = e^{\eta_i}$ (natural link function):

$$U_{ij} = \boldsymbol{x}_i'\left\{-\frac{1}{n}\sum_{k=1}^{n}\exp(\eta_k)\boldsymbol{x}_k\boldsymbol{x}_k'\right\}\boldsymbol{x}_j.$$

The CAIC obtained from the above coefficients coincides with that reported by Kamo *et al*. (2011).

- Case of the identity link function, i.e., $\lambda_i = \eta_i$:

$$c_{i1} = \frac{1}{\eta_i}, \quad c_{i2} = -\frac{1}{\eta_i^2}, \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{1}{n} \sum_{k=1}^{n} \frac{1}{\eta_k} \boldsymbol{x}_k \boldsymbol{x}_k' \right) \boldsymbol{x}_j.$$

- Case of the sqrt link function, i.e., $\lambda_i = \eta_i^2$:

$$c_{i1} = \frac{2}{\eta_i}, \quad c_{i2} = -\frac{2}{\eta_i^2}, \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{4}{n} \sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k' \right) \boldsymbol{x}_j.$$

4.3. Case of the Gamma Distribution

When we assume that a positive observed response, y_i, is distributed according to the Gamma distribution $\Gamma(\lambda_i, \nu)$ $(i = 1, \ldots, n)$, the parameters and functions based on the model are given as follows:

$$\begin{aligned} &\theta_i = -\frac{1}{\lambda_i \nu}, \quad b(\theta_i) = -\log(-\theta_i), \\ &\phi = \nu, \quad a(\phi) = \phi^{-1}, \quad c(y_i, \phi) = \nu \log \nu - \log(\Gamma(\nu)) + (\nu - 1) \log y_i, \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function. Then, the coefficients of the CAIC specifying the distribution are given by

$$d_{i1} = -\theta_i^{-1}, \quad d_{i2} = \theta_i^{-2}, \quad d_{i3} = -2\theta_i^{-3}, \quad d_{i4} = 6\theta_i^{-4}.$$

The remaining coefficients of the CAIC, which are determined by the link function, are as follows:

- Case of the inverse link function, i.e., $\mathrm{E}[y_i] = \lambda_i \nu = \eta_i^{-1}$ (natural link function):

$$U_{ij} = \boldsymbol{x}_i' \left(-\frac{\nu}{n} \sum_{k=1}^{n} \eta_k^{-2} \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

- Case of the log link function, i.e., $\lambda_i \nu = \exp(\eta_i)$:

$$c_{i1} = \exp(-\eta_i), \quad c_{i2} = -\exp(-\eta_i), \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{\nu}{n} \sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

- Case of the identity link function, i.e., $\lambda_i \nu = \eta_i$:

$$c_{i1} = \eta_i^{-2}, \quad c_{i2} = -2\eta_i^{-3} \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{\nu}{n} \sum_{k=1}^{n} \eta_k^{-4} \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

4.4. Case of the Inverse Gaussian Distribution

When we assume that a positive observed response, y_i, is distributed according to the inverse Gaussian distribution $IG(\mu_i, \lambda)$ $(i = 1, \ldots, n)$, the parameters and functions based on the model are given as follows:

$$\theta_i = \mu_i^{-2}, \quad b(\theta_i) = 2\sqrt{\theta_i},$$
$$\phi = \lambda, \quad a(\phi) = -\frac{\lambda}{2}, \quad c(y_i, \phi) = \frac{1}{2}\log(\lambda) - \frac{1}{2}\log(2\pi x^3) - \frac{\lambda}{2x},$$

Then, the coefficients of the CAIC specifying the distribution are given as

$$d_{i1} = \theta_i^{-1/2}, \quad d_{i2} = -\frac{1}{2}\theta_i^{-3/2}, \quad d_{i3} = 3\theta_i^{-5/2}, \quad d_{i4} = -\frac{15}{2}\theta_i^{-7/2}.$$

The remaining coefficients of the CAIC, which are determined by the link function, are as follows:

- Case of the $1/\mu^2$ link function, i.e., $\mathrm{E}[y_i] = \mu_i = \eta_i^{-1/2}$ (natural link function):

$$U_{ij} = \boldsymbol{x}_i' \left(-\frac{1}{n\lambda} \sum_{k=1}^{n} \eta_k^{-3/2} \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

- Case of the inverse link function, i.e., $\mu_i = \eta_i^{-1}$:

$$c_{i1} = 2\eta_i, \quad c_{i2} = 2, \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{4}{n\lambda} \sum_{k=1}^{n} \eta_k^{-1} \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

- Case of the log link function, i.e., $\mu_i = \exp(\eta_i)$:

$$c_{i1} = -2\exp(-2\eta_i), \quad c_{i2} = 4\exp(-2\eta_i), \quad U_{ij} = \boldsymbol{x}_i' \left(-\frac{4}{n\lambda} \sum_{k=1}^{n} \exp(-\eta_k) \boldsymbol{x}_k \boldsymbol{x}_k' \right)^{-1} \boldsymbol{x}_j.$$

5. NUMERICAL STUDIES

In this section, we conduct numerical studies to show that the CAIC is better than the original AIC. At the beginning of this section, we examine the numerical studies for the frequencies of the model and the prediction error of the best models selected by the criteria. We prepared the eight candidate models with $n = 50$ and 100. First, we constructed an $n \times 8$ explanatory variable matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$. The first column of X is $\mathbf{1}_n$, where $\mathbf{1}_n$ is an n-dimensional vector of ones, and the remaining seven columns of X were defined by realizations of independent dummy variables with binomial distribution $B(1, 0.4)$. In this simulation, we prepared two parameters $\boldsymbol{\beta}$, as follows:

$$\text{Case1} : \boldsymbol{\beta} = (0.65, -0.65)', \quad \text{Case2} : \boldsymbol{\beta} = (0.1, 0.1, 0.3, -0.5)'.$$

The explanatory variables matrix in the jth model consists of the first j columns of X $(j = 1, \ldots, 8)$. Thus, in case 1, the true model is the second model, and in case 2, the true model is the fourth model. We simulated 1,000 realizations of $\boldsymbol{y} = (y_1, \ldots, y_n)$ in the probit regression model, i.e., $y_i \overset{\text{i.d}}{\sim} B(1, p_i)$, where $p_i = \Phi(\boldsymbol{x}_i' \boldsymbol{\beta})$ $(i = 1, \ldots, n)$.

Table 1: Selection-probability and prediction error for the case of $\boldsymbol{\beta} = (0.65, -0.65)'$

n	Model	1	2	3	4	5	6	7	8	PE_B
	Risk	66.29	**64.54**	65.58	67.06	68.78	70.45	72.37	74.47	-
50	AIC	28.4	**44.8**	8.0	6.2	3.2	3.2	3.3	2.9	68.34
	CAIC	34.5	**48.6**	7.8	4.8.	1.6	1.4.	1.0	0.3	**66.94**
	Risk	130.81	**126.19**	127.31	128.52	129.67	130.98	132.27	133.64	-
100	AIC	11.3	**58.8**	11.6	6.8	4.1	3.5	1.1	2.8	128.95
	CAIC	12.0	**62.4**	10.8	6.3	3.2	2.8	1.1	1.4	**128.55**

Table 2: Selection-probability and prediction error for the case in which $\boldsymbol{\beta} = (0.1, 0.1, 0.3, -0.5)'$

n	Model	1	2	3	4	5	6	7	8	PE_B
	Risk	**70.17**	71.36	72.13	71.99	73.63	75.22	77.25	79.22	-
50	AIC	**47.2**	7.7	7.7	19.4	6.4	5.4	3.0	3.2	74.8
	CAIC	**55.2**	8.7	7.9	17.9	4.9	2.5	1.8	1.1	**73.79**
	Risk	139.54	140.23	140.09	**137.91**	138.94	140.14	141.25	142.45	-
100	AIC	27.4	4.0	8.2	**40.3**	8.6	5.1	3.6	2.8	140.42
	CAIC	29.8	4.7	8.8	**40.9**	7.1	4.0	3.0	1.7	**140.29**

Tables 1 and 2 list the following properties.

(1) selection-probability: the frequency of the model chosen by minimizing the information criterion.

(2) prediction error of the best model (PE_B): the risk function of the model selected by the information criterion as the best model, which is estimated as

$$\widehat{\text{PE}}_\text{B} = \frac{1}{1000} \sum_{i=1}^{1000} \text{E}_{\boldsymbol{y}^*}[-2\ell(\hat{\boldsymbol{\beta}}_{\text{B}_i}; \boldsymbol{y}^*)],$$

where $\boldsymbol{y}^*$ is a future observation, and $\hat{\boldsymbol{\beta}}_{\text{B}_i}$ is the value of $\hat{\boldsymbol{\beta}}$ of the selected model at the ith iteration.

In particular, PE_B is an important property because it is equivalent to the expected KL information between the true model and the best model selected by the criteria. In case 1 with $n = 50$ and 100 and in case 2 with $n = 100$, the model having the smallest risk function (referred to as the principle best model) coincides with the true model. On the other hand, in case 2 with $n = 50$, the principle best model became the first model rather than the fourth model, i.e., the true model did not conform with the principle best model. This means that using a model that is smaller than the true model is better for prediction in case 2 with $n = 50$. From Tables 1 and 2, we can see that the selection-probabilities and prediction errors of the CAIC were improved in all situations in comparison with the AIC. We simulated several other models and obtained similar results.

Table 3: Selection-probability and estimated prediction error

Selected model	AIC			CAIC		
	Logistic	Probit	Total	Logistic	Probit	Total
X Ray	0	0	0	0	5	5
X Ray, Acid	0	3	3	0	3	3
X Ray, Age	0	1	1	0	1	1
X Ray, Grade	0	1	1	1	3	4
X Ray, Stage	25	5	30	24	38	62
X Ray, Grade, Acid	0	4	4	0	1	1
X Ray, Stage, Acid	5	22	27	0	15	15
X Ray, Stage, Age	0	4	4	0	6	6
X Ray, Stage, Grade	0	2	2	0	2	2
X Ray, Grade, Age, Acid	2	0	2	1	0	1
X Ray, Stage, Age, Acid	0	17	17	0	0	0
X Ray, Stage, Grade, Acid	1	6	7	0	0	0
X Ray, Stage, Grade, Age	0	1	1	0	0	0
X Ray, Stage, Grade, Age, Acid	0	1	1	0	0	0
$\widehat{\mathrm{PE}}_{\mathrm{B}}$	146.5			**145.8**		

Next, for the purpose of analyzing the GLM, we consider the data reported in Brown (1980), who discussed an experiment in which 53 prostate cancer patients underwent surgery to examine their lymph nodes for evidence of cancer. The response variable is the number of patients with nodal involvement, and there were five predictor variables: X Ray, Stage, Age, Acid, and Grade. First, we assume that the response variable y_i is distributed according to $B(1, p_i)$ $(i = 1, \ldots, n)$. For the link function, we prepare two functions: the logistic link function and the Probit link function. In this analysis, we select the link functions and variables simultaneously. Table 3 shows the selection-probability of the model selected by minimizing the information criterion and the estimated prediction error of the best model selected by the information criterion. We divide the data into calibration sample data and validation sample data. The sample sizes of the calibration sample and the validation sample were 43 and 10, respectively. The best subset of the variables and the link function were selected by criteria derived from the calibration sample. The selection-probabilities were evaluated from only the calibration sample. The prediction errors were estimated as follows. Let $\boldsymbol{d}_j = (d_{1j}, \ldots, d_{nj})$ be an n-dimensional vector expressing a pattern to leave out 10 data at the jth iteration $j = 1, \ldots, 100)$, i.e., $d_{ij} = 1$ or 0 and $\sum_{i=1}^{n} d_{ij} = 10$. Moreover, we let $\hat{\boldsymbol{\beta}}_{\mathrm{B},[-\boldsymbol{d}_j]}$ denote

$\hat{\boldsymbol{\beta}}_{[-\boldsymbol{d}_j]}$ of $\boldsymbol{\beta}$ of the best model evaluated from the calibration sample, where $\hat{\boldsymbol{\beta}}_{[-\boldsymbol{d}_j]}$ is given as

$$\hat{\boldsymbol{\beta}}_{[-\boldsymbol{d}_j]} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{53} (1 - d_{ij}) \log f(y_i; \boldsymbol{\beta}).$$

Finally, the estimated $\mathrm{PE_B}$ is given as

$$\widehat{\mathrm{PE}}_{\mathrm{B}} = \frac{1}{100} \sum_{j=1}^{100} \frac{1}{10} \sum_{i=1}^{53} d_{ij} \left\{ -2 \log f(y_i; \hat{\boldsymbol{\beta}}_{\mathrm{B},[-\boldsymbol{d}_j]}) \right\}.$$

Table 3 indicates that the models selected by the AIC were spread over a wider area than those of the CAIC, although the model most selected by the AIC is the same as that selected by the CAIC. In particular, the selection probability of the model most selected by the CAIC is much higher than that selected by the AIC. The estimated prediction error of the CAIC was smaller than that of the AIC. Thus, the CAIC is thought to have improved the accuracy of the original AIC.

Consequently, from Tables 1, 2, and 3, we recommend the use of the CAIC rather than the AIC for selecting variables in the GLMs.

6. CONCLUSION

In the present paper, we derived a simple formula for the CAIC in the GLM. All of the coefficients in our formula are the first through fourth derivatives of the log-likelihood function. The GLM can express a number of statistical models by changing the distribution and the link function and can be easily fitted to the real data using the function "glm" in the "R" software, which implements several distributions and link functions. Hence, based on the real data analysis, the present result is useful for real data analysis. Moreover, the numerical studies revealed that the CAIC is better than the original AIC.

For example, we presented explicit forms of the CAIC in all of the models that are implemented in "glm". These explicit forms are thus confirmed to be useful in real data analysis. Even if a researcher wants to use the CAIC in a model that for which an example CAIC has not yet been derived, the researcher can easily obtain the CAIC using formula manipulation software.

In the present paper, we deal primarily with variable selection. However, in the simulation of the real data analysis, we also considered the selection of the link function. If we choose the link function by minimizing the original AIC, the optimal link function is determined only by maximizing the log-likelihood function. On the other hand, if we use the CAIC to select the link function, the optimal link function is not determined only by maximizing the log-likelihood function. Thus, using the CAIC will allow us to select an appropriate link function.

As mentioned above, we confirm that the results are useful and user friendly.

APPENDIX

A.1 Derivation of the Third Term of (3.4)

In order to calculate the moments of $\boldsymbol{b}_1$ and Z_2, we rewrite the third term of (3.4) using $\boldsymbol{b}_1$, Z_2, and M_3 as

$$\frac{2}{n}\mathrm{E}\left[\sum_{i=1}^{n}\frac{(y_i-d_{i1})}{a(\phi)}\left\{c_{i1}\boldsymbol{x}_i'\boldsymbol{b}_2+\frac{1}{2}c_{i2}(\boldsymbol{x}_i'\boldsymbol{b}_1)^2\right\}\right]=\frac{1}{\sqrt{n}}\mathrm{E}\left[\boldsymbol{b}_1'M_3(\boldsymbol{b}_1\otimes\boldsymbol{b}_1)\right]+\frac{3}{\sqrt{n}}\mathrm{E}\left[\boldsymbol{b}_1'Z_2\boldsymbol{b}_1\right], \qquad \text{(A.1)}$$

where c_{ij} and d_{ij} are defined in (2.5), and U_{ij} is defined in (3.8). Let $\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})$ be the characteristic function of the distribution of $\boldsymbol{b}_1$, defined as

$$\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})=\mathrm{E}[\exp(i\boldsymbol{t}'\boldsymbol{b}_1)]=\prod_{j=1}^{n}\mathrm{E}[\exp\{i(y_j-d_{j1})s_j\}],$$
$$s_j=-\frac{1}{\sqrt{n}a(\phi)}c_{j1}\boldsymbol{t}'M_2^{-1}\boldsymbol{x}_j,$$

where $\boldsymbol{t}=(t_1,\ldots,t_p)'$. Note that $\mathrm{E}[\exp\{i(y-\mu)s\}]$ is the characteristic function of $y-\mu$, which is expressed as

$$\mathrm{E}[\exp\{i(y-\mu)s\}]=\exp\left\{\frac{b(\theta+isa(\phi))-b(\theta)}{a(\phi)}-i\mu s\right\}.$$

Therefore, we have

$$\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})=\exp\left\{\sum_{j=1}^{n}\left(\frac{b(\theta_j+is_ja(\phi))-b(\theta_j)}{a(\phi)}-id_{j1}s_j\right)\right\}.$$

Based on the property of the random variable with mean zero, the third moment is equivalent to the third cumulant. Since $|s_j|=O(n^{-1/2})$, $\log\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})$ can be expanded as

$$\begin{aligned}\log\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})&=\sum_{j=1}^{n}\left\{\frac{b(\theta_j+is_ja(\phi))-b(\theta_j)}{a(\phi)}-id_{j1}s_j\right\}\\&=\frac{1}{a(\phi)}\sum_{j=1}^{n}\left\{\frac{1}{2}d_{j2}(is_ja(\phi))^2+\frac{1}{6}d_{j3}(is_ja(\phi))^3+\frac{1}{24}d_{j4}(is_ja(\phi))^4\right\}+O(n^{-3/2}).\end{aligned}$$

Thus, the third cumulant of $\boldsymbol{b}_1=(b_{11},\ldots,b_{1p})'$ is computed through the derivative of $\log\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})$, i.e.,

$$\begin{aligned}\mathrm{E}[b_{1\alpha_1}b_{1\alpha_2}b_{1\alpha_3}]&=i^{-3}\left.\frac{\partial^3\log\varphi_{\boldsymbol{b}_1}(\boldsymbol{t})}{\partial t_{\alpha_1}\partial t_{\alpha_2}\partial t_{\alpha_3}}\right|_{\boldsymbol{t}=\boldsymbol{0}}\\&=a(\phi)^2\sum_{j=1}^{n}d_{j3}\frac{\partial s_j}{\partial t_{\alpha_1}}\frac{\partial s_j}{\partial t_{\alpha_2}}\frac{\partial s_j}{\partial t_{\alpha_3}}+O(n^{-3/2}).\end{aligned}$$

Note that

$$\frac{\partial s_j}{\partial t_{\alpha_l}}=-\frac{1}{\sqrt{n}a(\phi)}c_{j1}\boldsymbol{e}_{\alpha_l}'M_2^{-1}\boldsymbol{x}_j,$$

where $\boldsymbol{e}_j$ is the p-dimensional vector, the jth element of which is 1 and the other elements of which are 0. Thus, using Equations (2.5) and (3.8), we obtain

$$\frac{1}{\sqrt{n}}\mathrm{E}[\boldsymbol{b}_1' M_3(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)] = \frac{1}{n^3 a(\phi)} \sum_{i,j}^{n} d_{j3} c_{j1}^3 (d_{i3} c_{i1}^3 + 3 d_{i2} c_{i1} c_{i2}) U_{ij}^3 + O(n^{-2}). \tag{A.2}$$

Let $\varphi_{\boldsymbol{b}_1, Z_2}(\boldsymbol{t}, T_1)$ denote the characteristic function of the joint distribution for $\boldsymbol{b}_1$ and Z_2 as

$$\varphi_{\boldsymbol{b}_1, Z_2}(\boldsymbol{t}, T_1) = \exp\left\{ \sum_{j=1}^{n} \frac{b(\theta_j + i v_j a(\phi)) - b(\theta_j)}{a(\phi)} - i d_{j1} v_j \right\},$$

where $T_1 = (t_{ij}^{(1)})$ $(i, j = 1, \ldots, p)$ and

$$v_j = \frac{1}{\sqrt{n} a(\phi)} (-c_{j1} \boldsymbol{t}' M_2^{-1} \boldsymbol{x}_j + c_{j2} \boldsymbol{x}_j' T_1 \boldsymbol{x}_j).$$

In the same manner as in the calculation of $\log \phi_{\boldsymbol{b}_1}(\boldsymbol{t})$, we have

$$\begin{aligned} \log \varphi_{\boldsymbol{b}_1, Z_2}(\boldsymbol{t}, T) &= \sum_{j=1}^{n} \left\{ \frac{b(\theta_j + i v_j a(\phi)) - b(\theta_j)}{a(\phi)} - i d_{j1} v_j \right\} \\ &= \frac{1}{a(\phi)} \sum_{j=1}^{n} \left\{ \frac{1}{2} d_{j2} (i v_j a(\phi))^2 + \frac{1}{6} d_{j3} (i v_j a(\phi))^3 + \frac{1}{24} d_{j4} (i v_j a(\phi))^4 \right\} + O(n^{-3/2}). \end{aligned}$$

Note that

$$\frac{\partial v_k}{\partial t_i} = -\frac{1}{\sqrt{n} a(\phi)} c_{k1} \boldsymbol{e}_i' M_2^{-1} \boldsymbol{x}_k, \quad \frac{\partial v_k}{\partial T_{ij}} = \frac{1}{\sqrt{n} a(\phi)} c_{k2} (\boldsymbol{e}_i' \boldsymbol{x}_k)(\boldsymbol{e}_j' \boldsymbol{x}_k).$$

Hence, we obtain

$$\frac{1}{\sqrt{n}}\mathrm{E}[\boldsymbol{b}_1 Z_2 \boldsymbol{b}_1] = \frac{1}{n^2 a(\phi)} \sum_{i=1}^{n} d_{i3} c_{i1}^2 c_{i2} U_{ii}^2 + O(n^{-2}). \tag{A.3}$$

Substituting (A.2) and (A.3) into (A.1), the third term of (3.4) is given by (3.7).

A.2 Derivation of the Fourth Term of (3.4)

In order to use the asymptotic properties, we express the fourth term of (3.4) in terms of $\boldsymbol{b}_1$, Z_2, Z_3 M_2 M_3, and M_4 as

$$\begin{aligned} &\frac{2}{n\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^{n} \frac{(y_i - d_{i1})}{a(\phi)} \left\{ c_{i1} \boldsymbol{x}_i' \boldsymbol{b}_3 + c_{i2} (\boldsymbol{x}_i' \boldsymbol{b}_1)(\boldsymbol{x}_i' \boldsymbol{b}_2) + \frac{1}{6} c_{i3} (\boldsymbol{x}_i' \boldsymbol{b}_1)^3 \right\} \right] \\ &= -\frac{1}{n}\mathrm{E}\left[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) \right] + \frac{1}{3n}\mathrm{E}\left[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_4 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) \right] \\ &\quad - \frac{3}{n}\mathrm{E}\left[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} Z_2 \boldsymbol{b}_1 \right] - \frac{4}{n}\mathrm{E}\left[\boldsymbol{b}_1' Z_2 M_2^{-1} Z_2 \boldsymbol{b}_1 \right] + \frac{4}{3n}\mathrm{E}\left[\boldsymbol{b}_1' Z_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1) \right]. \end{aligned} \tag{A.4}$$

Let $\varphi_{\boldsymbol{b}_1,Z_2,Z_3}(\cdot)$ denote the characteristic function of the joint distribution for $\boldsymbol{b}_1$, Z_2, and Z_3 defined by

$$\varphi_{\boldsymbol{b}_1,Z_2,Z_3}(\boldsymbol{t},T_2,T_3)=\exp\left\{\sum_{j=1}^{n}\left(\frac{b(\theta_j+ir_ja(\phi))-b(\theta_j)}{a(\phi)}-id_{j1}r_j\right)\right\},$$

where $T_2=(T_{ij}^{(2)})$ $(i,j=1,\dots,p)$, $T_3=(T_{ijk}^{(3)})$ $(i,j,k=1,\dots,p)$ and

$$r_j=\frac{1}{\sqrt{n}a(\phi)}(-c_{j1}\boldsymbol{t}'M_2^{-1}\boldsymbol{x}_j+c_{j2}\boldsymbol{x}_j'T_2\boldsymbol{x}_j+c_{j3}\boldsymbol{x}_j'T_3(\boldsymbol{x}_j\otimes\boldsymbol{x}_j)).$$

In order to simplify the calculations, we define the following notation:

$$\begin{aligned}
\tau_{kj}&=\frac{1}{a(\phi)}\frac{1}{k!}(ia(\phi)d_{jk}r_j)^k,\\
\kappa_{ij}&=\sum_{\alpha=1}^{n}\frac{\partial^2\tau_{2\alpha}}{\partial t_i\partial t_j}=-\frac{1}{na(\phi)}\sum_{m=1}^{n}d_{m2}c_{m1}^2(\boldsymbol{e}_i'M_2^{-1}\boldsymbol{x}_m)(\boldsymbol{e}_j'M_2^{-1}\boldsymbol{x}_m), &\text{(A.5)}\\
\kappa_{i,kl}&=\sum_{\alpha=1}^{n}\frac{\partial^2\tau_{2\alpha}}{\partial t_i\partial T_{jk}^{(2)}}=\frac{1}{na(\phi)}\sum_{m=1}^{n}d_{m2}c_{m1}c_{m2}(\boldsymbol{e}_i'M_2^{-1}\boldsymbol{x}_m)(\boldsymbol{e}_k'\boldsymbol{x}_m)(\boldsymbol{e}_l'\boldsymbol{x}_m), &\text{(A.6)}\\
\kappa_{ik,jl}&=\sum_{\alpha=1}^{n}\frac{\partial^2\tau_{2\alpha}}{\partial T_{ik}^{(2)}\partial T_{jl}^{(2)}}=-\frac{1}{na(\phi)}\sum_{m=1}^{n}d_{m2}c_{m2}^2(\boldsymbol{e}_i'\boldsymbol{x}_m)(\boldsymbol{e}_k'\boldsymbol{x}_m)(\boldsymbol{e}_j'\boldsymbol{x}_m)(\boldsymbol{e}_l'\boldsymbol{x}_m), &\text{(A.7)}\\
\kappa_{i,ijk}&=\sum_{\alpha=1}^{n}\frac{\partial^2\tau_{2\alpha}}{\partial t_i\partial T_{ijk}^{(3)}}=\frac{1}{na(\phi)}\sum_{m=1}^{n}d_{m2}c_{m1}c_{m3}(\boldsymbol{e}_i'M_2^{-1}\boldsymbol{x}_m)(\boldsymbol{e}_i'\boldsymbol{x}_m)(\boldsymbol{e}_j'\boldsymbol{x}_m)(\boldsymbol{e}_k'\boldsymbol{x}_m). &\text{(A.8)}
\end{aligned}$$

Using the derivations of $\phi_{\boldsymbol{b}_1,Z_2,Z_3}$, the first term of (A.4) is given by

$$\begin{aligned}
\mathrm{E}[(\boldsymbol{b}_1\otimes\boldsymbol{b}_1)'M_3'M_2^{-1}M_3(\boldsymbol{b}_1\otimes\boldsymbol{b}_1)]&=\sum_{i,j,k,l}^{p}[M_3'M_2^{-1}M_3]_{i,j,k,l}\mathrm{E}[b_{i1}b_{j1}b_{k1}b_{l1}]\\
&=\sum_{i,j,k,l}^{p}[M_3'M_2^{-1}M_3]_{i,j,k,l}\left.\frac{\partial^4\varphi_{\boldsymbol{b}_1,Z_2,Z_3}(\boldsymbol{t},T_2,T_3)}{\partial t_i\partial t_j\partial t_k\partial t_l}\right|_{\boldsymbol{t}=\boldsymbol{0}_p,T_2=0,T_3=0}.
\end{aligned}$$

By applying a Taylor expansion, we obtain

$$\begin{aligned}
&\frac{\partial^4}{\partial t_i\partial t_j\partial t_k\partial t_l}\exp\left\{\sum_{j=1}^{n}\left(\frac{b(\theta_j+ir_ja(\phi))-b(\theta_j)}{a(\phi)}-id_{j1}r_j\right)\right\}\Bigg|_{\boldsymbol{t}=\boldsymbol{0}_p,T_2=0,T_3=0}\\
&=\frac{\partial^4}{\partial t_i\partial t_j\partial t_k\partial t_l}\exp\left\{\sum_{j=1}^{n}(\tau_{2j}+\tau_{3j}+\tau_{4j})+O(n^{-3/2})\right\}\Bigg|_{\boldsymbol{t}=\boldsymbol{0}_p,T_2=0,T_3=0}\\
&=\left\{\kappa_{ij}\kappa_{kl}+\kappa_{ik}\kappa_{jl}+\kappa_{jk}\kappa_{il}+\sum_{\alpha=1}^{n}\frac{\partial^4\tau_{4\alpha}}{\partial t_i\partial t_j\partial t_k\partial t_l}+O(n^{-2/3})\right\}\exp\left\{1+O(n^{-3/2})\right\}.
\end{aligned}$$

Note that $|r_j|=O(n^{-1/2})$ and

$$\frac{\partial^4\tau_{4\alpha}}{\partial t_i\partial t_j\partial t_k\partial t_l}=\sum_{\alpha=1}^{n}a(\phi)^3d_{\alpha4}\frac{\partial r_\alpha}{\partial t_i}\frac{\partial r_\alpha}{\partial t_j}\frac{\partial r_\alpha}{\partial t_k}\frac{\partial r_\alpha}{\partial t_l}=O(n^{-1}).$$

Hence, the first term of (A.4) is expressed as

$$\mathrm{E}[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)] = \sum_{i,j,k,l}^{p} [M_3' M_2^{-1} M_3]_{i,j,k,l} (\kappa_{ij}\kappa_{kl} + \kappa_{ik}\kappa_{jl} + \kappa_{jk}\kappa_{il}) + O(n^{-1}). \quad \text{(A.9)}$$

By substituting (A.5) into (A.9), we obtain

$$\begin{aligned} &\mathrm{E}[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} M_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)] \\ &= \frac{1}{n^2 a(\phi)^2} \sum_{i,j}^{n} (d_{i2} c_{i1}^3 + 3 d_{i2} c_{i1} c_{i2})(d_{j2} c_{j1}^3 + 3 d_{j2} c_{j1} c_{j2})(U_{ii} U_{ij} U_{jj} + 2 U_{ij}^3) + O(n^{-1}). \end{aligned} \quad \text{(A.10)}$$

The remaining terms of (A.4), as well as the first term of (A.4), will be calculated. The second term of (A.4) is similarly obtained from (A.10) as follows:

$$\mathrm{E}[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_4 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)] = -\frac{3}{n} \frac{1}{a(\phi)} \sum_{i=1}^{n} (d_{i4} c_{i1}^4 + 6 d_{i3} c_{i1}^2 c_{i2} + 3 d_{i2} c_{i2}^2 + 4 d_{i2} c_{i1} c_{i3}) U_{ii}^2 + O(n^{-1}). \quad \text{(A.11)}$$

Next, we calculate the third term of (A.4). The third term of (A.4) is expressed as follows:

$$\begin{aligned} \mathrm{E}\left[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} Z_2 \boldsymbol{b}_1\right] &= \sum_{i,j,k,l}^{p} [M_3' M_2^{-1}]_{i,j,k} \mathrm{E}[b_{1i} b_{1j} b_{1l} Z_{2,kl}] \\ &= \sum_{i,j,k,l}^{p} [M_3' M_2^{-1}]_{i,j,k} (\kappa_{ij}\kappa_{l,kl} + \kappa_{ik}\kappa_{j,kl} + \kappa_{jk}\kappa_{i,kl}) + O(n^{-1}), \end{aligned} \quad \text{(A.12)}$$

Expression (A.6) implies that

$$\begin{aligned} &\mathrm{E}\left[(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)' M_3' M_2^{-1} Z_2 \boldsymbol{b}_1\right] \\ &= -\frac{1}{n^2 a(\phi)^2} \sum_{i,j}^{n} (d_{i3} c_{i1}^3 + 3 d_{i2} c_{i1} c_{i2}) d_{j2} c_{j1} c_{j2} (U_{ii} U_{ij} U_{jj} + 2 U_{ij}^3) + O(n^{-1}). \end{aligned} \quad \text{(A.13)}$$

The fourth term of (A.4) is as follows:

$$\mathrm{E}\left[\boldsymbol{b}_1' Z_2 M_2^{-1} Z_2 \boldsymbol{b}_1\right] = \sum_{i,j,k,l}^{p} [M_2^{-1}]_{jk} (\kappa_{il}\kappa_{ik,jl} + \kappa_{i,ij}\kappa_{l,kl} + \kappa_{i,kl}\kappa_{l,ij}) + O(n^{-1}). \quad \text{(A.14)}$$

It follows from (A.7) and (A.14) that

$$\begin{aligned} &\mathrm{E}\left[\boldsymbol{b}_1' Z_2 M_2^{-1} Z_2 \boldsymbol{b}_1\right] \\ &= -\frac{1}{n a(\phi)} \sum_{i=1}^{n} d_{i2} c_{i2}^2 U_{ii} + \frac{1}{n^2 a(\phi)^2} \sum_{i,j}^{n} (d_{i2} c_{i1} c_{i2})(d_{j2} c_{j1} c_{j2})(U_{ii} U_{ij} U_{jj} + U_{ij}^3) + O(n^{-1}). \end{aligned} \quad \text{(A.15)}$$

Finally, we calculate the fifth term of (A.4). Note that

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{b}_1' Z_3 (\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)\right] &= \sum_{i,j,k}^{p} \mathrm{E}[Z_{3,ijk} b_{1i} b_{1j} b_{1k}] \\ &= \sum_{i,j,k}^{p} (\kappa_{ij}\kappa_{k,ijk} + \kappa_{ik}\kappa_{j,ijk} + \kappa_{jk}\kappa_{i,ijk}) + O(n^{-1}). \end{aligned} \quad \text{(A.16)}$$

Substituting (A.8) into (A.16) yields

$$\mathrm{E}\left[\boldsymbol{b}_1' Z_3(\boldsymbol{b}_1 \otimes \boldsymbol{b}_1)\right] = \frac{3}{n}\sum_{i=1}^{n} d_{i2}c_{i1}c_{i3}U_{ii}^2 + O(n^{-1}). \tag{A.17}$$

Consequently, from (A.10), (A.11), (A.13), (A.15), and (A.17), we obtain the fourth term of (3.4) as (3.9).

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Mariko Yamamura of Hiroshima University and Dr. Eisuke Inoue of Kitazato University for their helpful comments and suggestions. The authors would also like to thank Dr. Kenichi Kamo of Sapporo Medical University for teaching us the R programming langauge.

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (eds. Petrov, B. N. & Csáki, F.), 267–281, Akadémiai Kiadó, Budapest.

[2] Barnett, S. B. L. & Nurmagambetov, T. A. (2010). Cost of asthma in the United States: 2002-2007. *J. Allergy Clin. Immun.*, **127**, 145–152.

[3] Brown, B. W. (1980). Prediction analysis for binary data. In *Biostatistics Casebook.* (eds. Miller Jr., G. R., Efron., B., Brown, B. W., & Moses, L. E.), 3–18, Wiley, New York.

[4] Davies, S. L., Neath, A. A. & Cavanaugh, J. E. (2006). Estimation optimality of corrected AIC and modified *Cp* in linear regression. *Internat. Statist. Rev.*, **74**, 2, 161–168.

[5] Fahrmeir, L. & Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, **13**, 342–368.

[6] Hurvich, C. M. & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.

[7] Kamo, K., Yanagihara, H. & Satoh, K. (2011). Bias-corrected AIC for selecting variables in Poisson regression models. *Comm. Statist. Theory Methods* (in press).

[8] Kullback, S. & Libler, R. (1951). On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.

[9] Nelder, J. A. & Wedderburn, W. M. (1972). Generalized linear models. *J. R. Statist. Soc. A.*, **135**, 370–384.

[10] Matas, M., Picornell, A., Cifuentes, C., Payeras, A., Bassa, A., Homar, F., López-Labrador, F. X., Moya, A., Ramon, M. M. & Castro, J. A. (2010). Relating the liver damage with hepatitis C virus polymorphism in core region and human variables in HIV-1-coinfected patients. *Infect. Genet. Evol.*, **10**, 1252–1261.

[11] Meyers, R. H., Montgomery, D. C. & Vining, G. G. (2002). *Generalized Linear Models with Applications in Engineering and the Sciences.* Wiley Interscience, Canada.

[12] R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[13] Sánchez-Carneo, N., Couñago E., Rodrigues-Perez, D. & Freire, J. (2011). Exploiting oceanographic satellite data to study the small scale coastal dynamics in a NE Atlantic open embayment. *J. Marine Syst.*, **87**, 123–132.

[14] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Math. Statist.*, **8**, 147–164.

[15] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Statist. -Theory Meth.*, **7**, 1, 13–26.

[16] Teste, F. P. & Lieffers, V. J. (2011). Snow damage in lodgepole pine stands brought into thinning and fertilization regimes. *Forest Ecol. Manag.*, **261**, 2094–2104.

[17] Wong, C. S. & Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models. *J. Time Ser. Anal.*, **19**, 113–124.

[18] Yanagihara, H., Sekiguchi, R. & Fujikoshi, Y. (2003). Bias correction of AIC in logistic regression models. *J. Statist. Plann. Inference*, **115**, 349–360.