

Bootstrap Information Criteria for Linear Mixed Models

Yiping, Tang

Graduate School of Science, Chiba University

1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

E-mail: tang@graduate.chiba-u.jp

Jinfang, Wang

Department of Mathematics and Informatics, Chiba University

1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

E-mail: wang@math.s.chiba-u.ac.jp

1 Introduction

Linear mixed models (LMM) have received tremendous attention in the literature since the seminal paper by Laird and Ware (1982) due to their ability to represent clustered (therefore dependent) data. Model selection in LMM have traditionally been carried out using the marginal version of the Akaike information criterion (mAIC), which is designed for cross-sectional data (Pinheiro and Bates (2000); Ngo and Brand (2002)). In an important paper, Vaida and Blanchard (2005) pointed out that when the focus of the research is on the clusters instead of the population a more appropriate criterion for selecting LMM is to use the conditional version of the Akaike information criterion (cAIC). Since Vaida and Blanchard (2005) a number of papers have been published recently discussing the use of mAIC and cAIC (see e.g. Liang, et al. (2008), Greven and Kneib (2010) and Srivastava and Kubokawa (2010)). In this section we give a review of mAIC and the basic version of cAIC proposed originally by Vaida and Blanchard (2005), which seems to be used in practice (Greven and Kneib, 2010).

Suppose that we have data from m clusters, in the i th cluster the n_i -vector of response \mathbf{y}_i being modeled by $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, i = 1, \dots, m$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a vector of fixed effects, $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})'$ is a vector of random effects, \mathbf{X}_i and \mathbf{Z}_i are the $n_i \times p$ and $n_i \times q$ covariate matrices for the fixed and random effects of full column ranks. The error vectors $\boldsymbol{\epsilon}_i$ and the random effects \mathbf{b}_i are independently and normally distributed, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$, $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \sigma^2 \mathbf{K})$, where \mathbf{I}_{n_i} is the $n_i \times n_i$ identity matrix and \mathbf{K} a $q \times q$ positive definite matrix. It is possible to write the linear mixed model (LMM) in a single equation by stacking the vectors and matrices appropriately. Specifically, let $n = \sum_{i=1}^m n_i$ be the total number of observations, $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$ the n -vector of observations, $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_m)'$ the $n \times p$ covariate matrix for the fixed effects, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$ the $n \times mq$ block diagonal covariate matrix for the random effects, $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_m)'$ the mq -vector of random effects, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_1, \dots, \boldsymbol{\epsilon}'_m)'$ the n -vector of errors, and $\sigma^2 \mathbf{D} = \sigma^2 \text{diag}(\mathbf{K}, \dots, \mathbf{K})$ the $mq \times mq$ block diagonal covariance matrix of \mathbf{b} . Then LMM can be written as

$$(1) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad \begin{pmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim \mathcal{N} \left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_n \end{pmatrix} \right\}.$$

Let $f_c(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \sigma^2, \mathbf{K}) = (2\pi\sigma^2)^{-n/2} \exp\{-|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}|^2/(2\sigma^2)\}$ denote the conditional density function of \mathbf{y} given \mathbf{b} . Let $p(\mathbf{b}|\mathbf{K}) = \prod_{j=1}^m p(\mathbf{b}_j|\mathbf{K})$ denote the joint density function of \mathbf{b} . The marginal likelihood $f_m(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{K})$ of \mathbf{y} can be written as

$$(2) \quad f_m(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{K}) = \frac{1}{(2\pi\sigma^2)^{n/2} |\mathbf{V}|^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\},$$

where $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{I}_n$. Suppose that \mathbf{K} (therefore \mathbf{V}) is known. Then the maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 can be obtained in a straightforward manner:

$$(3) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad \hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / n.$$

Putting β and $\hat{\sigma}^2$ into (20) we have $f_m(\mathbf{y}|\beta, \hat{\sigma}^2, \mathbf{K}) = (2\pi\hat{\sigma}^2)^{-n/2}|\mathbf{V}|^{-1/2}e^{-n/2}$. So the marginal Akaike information criterion and the finite sample corrected version take the following respective forms

$$(4) \quad \text{mAIC} = n \log(2\pi\hat{\sigma}^2) + n + \log|\mathbf{V}| + 2(p + 1),$$

$$(5) \quad \text{mAIC}_c = n \log(2\pi\hat{\sigma}^2) + n + \log|\mathbf{V}| + 2n(p + 1)/(n - p - 2),$$

where p is the dimension of β . In practice \mathbf{K} is often unknown. In this case we may still use (4) and (5) but with \mathbf{K} replaced by a consistent estimator $\widehat{\mathbf{K}}$ (Srivastava and Kubokawa, 2010).

The marginal Akaike information criteria mAIC and mAIC_c can be viewed as estimators for the marginal Akaike information which is defined as

$$(6) \quad \begin{aligned} \text{mAI} &= -2E_{G(\mathbf{y})}E_{G(\mathbf{w})} \log f_m(\mathbf{w}|\hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \\ &= -2 \int \int \log f_m(\mathbf{w}|\hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) f_m(\mathbf{w}|\beta, \sigma^2, \widehat{\mathbf{K}}) f_m(\mathbf{y}|\beta, \sigma^2, \widehat{\mathbf{K}}) d\mathbf{w}d\mathbf{y}. \end{aligned}$$

Vaida and Blanchard (2005) argue that when the focus of the research is on the clusters rather than on the population, a more appropriate information measure should be defined in a conditional manner. They define the conditional Akaike information as

$$(7) \quad \begin{aligned} \text{cAI} &= -2E_{G(\mathbf{y}, \mathbf{b})}E_{G(\mathbf{w}|\mathbf{b})} \log f_c(\mathbf{w}|\hat{\mathbf{b}}, \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \\ &= -2 \int \int \log f_c(\mathbf{w}|\hat{\mathbf{b}}, \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) g(\mathbf{w}|\mathbf{b})g(\mathbf{y}, \mathbf{b}) d\mathbf{w}d\mathbf{y}d\mathbf{b}, \end{aligned}$$

where $g(\mathbf{y}, \mathbf{b}) = g(\mathbf{y}|\mathbf{b})p(\mathbf{b})$ is the true joint density function of \mathbf{y} and \mathbf{b} , $g(\mathbf{w}|\mathbf{b})$ is the conditional density of an independent future observation w which shares the same random effect \mathbf{b} with \mathbf{y} with distribution $p(\mathbf{b})$. In principle $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\mathbf{b}}$ in (7) can be any estimators. Popular choices for $\hat{\beta}$ and $\hat{\sigma}^2$ are the maximum likelihood estimators given by (3), and $\hat{\mathbf{b}}$ is the empirical Bayes estimator given by

$$(8) \quad \hat{\mathbf{b}} = E(\mathbf{b}|\hat{\beta}, \hat{\sigma}^2, \mathbf{y}) = \mathbf{DZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}).$$

When the σ^2 and \mathbf{K} are known, an asymptotically unbiased estimator of cAI is given by

$$(9) \quad \text{cAIC} = -2 \log f_c(\mathbf{y}|\hat{\mathbf{b}}, \hat{\beta}, \sigma^2, \widehat{\mathbf{K}}) + 2(\rho + 1),$$

where $\hat{\mathbf{b}}$ is the empirical Bayes estimator in (8), $\hat{\beta}$ is the maximum likelihood estimator, and $\rho = \text{tr}(\mathbf{H}_1)$ is the effective number of degrees of freedom for model (20) with the form of

$$\rho = \text{tr} \left\{ \begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{D}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} \end{pmatrix} \right\}.$$

See Vaida and Blanchard (2005).

2 Bootstrap Information Criterion

Ishiguro et al. (1997) have proposed an extension of AIC based on the bootstrap. They call it the extended information criterion (EIC). A salient advantage of the bootstrap method is that it uses massive iterative computer calculations rather than analytic expressions, so it is free from troublesome analytic derivation of the bias correction term; moreover, it can be applied to almost any type of models and estimation procedures under very weak assumption. In the bootstrap methods, the true distribution $G(\mathbf{y})$ is substituted by an empirical distribution function $\widehat{G}(\mathbf{y})$. A random sample from $\widehat{G}(\mathbf{y})$ is called bootstrap sample, and is denoted as $\mathbf{y}^* = \{y_1^*, \dots, y_N^*\}$.

A statistical model $f(\mathbf{y}|\hat{\boldsymbol{\theta}}^*)$ is constructed based on the bootstrap sample \mathbf{y}^* , with $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}(\mathbf{y}^*)$. The bootstrap version of the expected log-likelihood can be rewritten as

$$(10) \quad E_{\widehat{G}(\mathbf{w})}[\log f(\mathbf{w}|\hat{\boldsymbol{\theta}}^*)] = \int \log f(\mathbf{w}|\hat{\boldsymbol{\theta}}^*)d\widehat{G}(\mathbf{w}) = \frac{1}{N} \sum_{j=1}^N \log f(y_j|\hat{\boldsymbol{\theta}}^*) \equiv \ell_N(\mathbf{y}|\hat{\boldsymbol{\theta}}^*).$$

The bootstrap estimator of the expected log-likelihood is defined as follows:

$$(11) \quad E_{\widehat{G}^*(w)}[\log f(W|\widehat{\theta}^*)] = \int \log f(w|\widehat{\theta}^*)d\widehat{G}^*(w) = \frac{1}{N} \sum_{j=1}^N \log f(y_j^*|\widehat{\theta}^*) \equiv \ell_N(\mathbf{y}^*|\widehat{\theta}^*),$$

where $\widehat{G}^*(w)$ is the empirical distribution function based on \mathbf{y}^* . The bootstrap bias estimate is

$$(12) \quad b^* = NE_{\widehat{G}(\mathbf{y}^*)} \left[\ell_N(\mathbf{y}^*|\widehat{\theta}^*) - \ell_N(\mathbf{y}|\widehat{\theta}^*) \right].$$

Extract B sets of bootstrap samples of size N , and write the i th bootstrap sample as $\mathbf{y}^*(i) = \{y_1^*(i), \dots, y_N^*(i)\}$. The bias estimate in (12) is usually numerically approximated by

$$(13) \quad b^* \approx \frac{1}{B} \sum_{i=1}^B \left[\ell_N(\mathbf{y}^*(i)|\widehat{\theta}^*(i)) - \ell_N(\mathbf{y}(i)|\widehat{\theta}^*(i)) \right] \equiv b_B,$$

where $\widehat{\theta}^*(i)$ is an estimate of θ using $\mathbf{y}^*(i)$. The extended information criterion is defined as

$$(14) \quad \text{EIC} = -2 \left(N\ell_N(\mathbf{y}|\widehat{\theta}) - b^* \right).$$

Konishi and Kitagawa (1996) have given a theoretical justification of EIC.

2.1 Variance Reduction in Bootstrap Simulation

There are two kinds of errors in the bootstrap bias estimate b_B . One is caused by the randomness of the observed data, the other is the simulation error which decreases as the number of bootstrap replication increases. Konishi and Kitagawa (1996, 2008, 2010) considered an efficient resampling method to reduce the second type of error. Let $\mathbf{T}(\cdot) = (T(\cdot), \dots, T_p(\cdot))^T \in \mathbf{R}^p$ be a p -dimensional functional estimator. Then the difference between (10) and (11) can be decomposed into three terms as follows:

$$(15) \quad D(\mathbf{y}; G) = \left[\ell_N(\mathbf{y}|\widehat{\theta}) - \int \log f(w|\widehat{\theta})dG(w) \right] = D_1(\mathbf{y}; G) + D_2(\mathbf{y}; G) + D_3(\mathbf{y}; G),$$

where

$$(16) \quad \begin{aligned} D_1(\mathbf{y}; G) &= \ell_N(\mathbf{y}|\widehat{\theta}) - \ell_N(\mathbf{y}|\mathbf{T}(G)), \\ D_2(\mathbf{y}; G) &= \ell_N(\mathbf{y}|\mathbf{T}(G)) - \int \log f(w|\mathbf{T}(G))dG(w), \\ D_3(\mathbf{y}; G) &= \int \log f(w|\mathbf{T}(G))dG(w) - \int \log f(w|\widehat{\theta})dG(w), \end{aligned}$$

with $\widehat{\theta}$ being the functional estimator such that $\widehat{\theta} = \mathbf{T}(\widehat{G})$.

By taking the expectation and the variance term by term on the right-hand side of (16), we observe that the expectation of $D_2(\mathbf{y}; G)$ is zero, so $E_{G(\mathbf{y})} [D(\mathbf{y}; G)] = E_{G(\mathbf{y})} [D_1(\mathbf{y}; G) + D_3(\mathbf{y}; G)]$, and $Var [D(\mathbf{y}; G)] = O(n^{-1})$ and in contrast $Var [D_1(\mathbf{y}; G) + D_3(\mathbf{y}; G)] = O(n^{-2})$. Therefore, for the bootstrap estimate we have

$$(17) \quad E_{\widehat{G}(\mathbf{y}^*)} \left[D(\mathbf{y}^*; \widehat{G}) \right] = E_{\widehat{G}(\mathbf{y}^*)} \left[D_1(\mathbf{y}^*; \widehat{G}) + D_3(\mathbf{y}^*; \widehat{G}) \right],$$

where $D_1(\mathbf{y}^*; \widehat{G}) = \ell_N(\mathbf{y}^*|\widehat{\theta}^*) - \ell_N(\mathbf{y}^*|\widehat{\theta})$ and $D_3(\mathbf{y}^*; \widehat{G}) = \ell_N(\mathbf{y}|\widehat{\theta}) - \ell_N(\mathbf{y}|\widehat{\theta}^*)$.

To reduce the fluctuation in the bootstrap bias estimation of log-likelihood, we use the following formula as a bootstrap bias estimate:

$$(18) \quad b_B = \frac{1}{B} \sum_{i=1}^B \left\{ D_1(\mathbf{y}^*(i); \widehat{G}) + D_3(\mathbf{y}^*(i); \widehat{G}) \right\}.$$

3 Bootstrap Information Criteria for the Linear Mixed Models

3.1 Marginal extended information criterion: mEIC

Since the maximized marginal likelihood can be calculated straightforwardly by putting $\hat{\beta}$ and $\hat{\sigma}^2$ into (20), we focus on the construction of the bootstrap bias estimator. Because \mathbf{K} is often unknown, we just consider the case for unknown \mathbf{K} . By using the maximized marginal likelihood, the bias of the marginal likelihood is given by:

$$(19) \quad b_m = E_{G(\mathbf{y})} \left[\sum_{j=1}^m \left[\log f_m(\mathbf{y}_j | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right] - E_{G(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right] \right]$$

In applying the bootstrap method, the second term in b_m , namely the expected log-likelihood $E_{G(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right]$ is replaced by $E_{\widehat{G}(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right]$, where \mathbf{w} is an n -dimensional future variable sharing the same random effects with \mathbf{y} , and the density function of $\widehat{G}(\mathbf{w})$ is given by

$$(20) \quad g_m(\mathbf{w} | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2} |\widehat{\mathbf{V}}|^{1/2}} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} (\mathbf{w} - \mathbf{X}\hat{\beta})' \widehat{\mathbf{V}}^{-1} (\mathbf{w} - \mathbf{X}\hat{\beta}) \right\},$$

Further the expectation $E_{G(\mathbf{y})}$ is replaced by the bootstrap expectation $E_{\widehat{G}(\mathbf{y}^*)}$ of the bootstrap sample \mathbf{y}^* , which is defined as follows

$$(21) \quad \mathbf{y}^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{b}^* + \boldsymbol{\epsilon}^*, \quad \begin{pmatrix} \mathbf{b}^* \\ \boldsymbol{\epsilon}^* \end{pmatrix} \sim \mathcal{N} \left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2 \widehat{\mathbf{D}} & \mathbf{0} \\ \mathbf{0} & \hat{\sigma}^2 \mathbf{I}_n \end{pmatrix} \right\}.$$

Consequently, the bootstrap estimate of b_m is given by

$$(22) \quad b_m^* = E_{\widehat{G}(\mathbf{y}^*)} \left[\log f_m(\mathbf{y}^* | \hat{\beta}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) - E_{\widehat{G}(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right] \right],$$

where $\hat{\beta}^* = \hat{\beta}(\mathbf{y}^*)$, $\hat{\sigma}^{2*} = \hat{\sigma}(\mathbf{y}^*)^2$, $\widehat{\mathbf{K}}^* = \widehat{\mathbf{K}}(\mathbf{y}^*)$. Using this bootstrap estimate of bias, we define the bootstrap version of the marginal information criterion in LMM as follows:

$$(23) \quad \text{mEIC} = -2 \left(\log f_m(\mathbf{y} | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) - b_m^* \right).$$

Furthermore, let us use the variance reduction method to reduce the fluctuation in the bootstrap bias estimation of marginal log-likelihood.

$$(24) \quad b_m^* \approx \frac{1}{B} \sum_{i=1}^B \{ D_{m1}(\mathbf{y}^*(i)) + D_{m3}(\mathbf{y}^*(i)) \},$$

where

$$\begin{aligned} D_{m1}(\mathbf{y}^*(i)) &= \log f_m(\mathbf{y}^* | \hat{\beta}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) - \log f_m(\mathbf{y}^* | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}), \\ D_{m3}(\mathbf{y}^*(i)) &= E_{\widehat{G}(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right] - E_{\widehat{G}(\mathbf{w})} \left[\log f_m(\mathbf{w} | \hat{\beta}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right]. \end{aligned}$$

3.2 Conditional extended information criterion: cEIC

When we consider the bootstrap bias estimator for the conditional likelihood, in addition to the estimators $\hat{\beta}$, $\hat{\sigma}^2$ and $\widehat{\mathbf{K}}$, we also need to estimate the random effect \mathbf{b} , which is usually estimated by the empirical Bayes method. The procedure to construct a conditional bootstrap information criterion in LMM is similar to that for constructing mEIC. We use the maximized conditional log-likelihood to get the bias of the conditional likelihood as follows:

$$b_c = E_{G(\mathbf{y}, \mathbf{b})} \left[\log f_c(\mathbf{y} | \hat{\mathbf{b}}, \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) - E_{G(\mathbf{w} | \mathbf{b})} \left[\log f_c(\mathbf{w} | \hat{\mathbf{b}}, \hat{\beta}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right] \right].$$

In applying the bootstrap method, similar to the case in constructing the marginal information criterion, the expected conditional log-likelihood, $E_{G(\mathbf{w}|\mathbf{b})} \left[\log f_c(\mathbf{w}|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right]$ is replaced by the conditional expectation $E_{\widehat{G}_c} \left[\log f_c(\mathbf{w}|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right]$, where \widehat{G}_c has density function

$$(25) \quad g_c(\mathbf{w}|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{K}) = (2\pi\hat{\sigma}^2)^{-n/2} \exp \left\{ -|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}|^2 / (2\hat{\sigma}^2) \right\}$$

In the definition of b_c , the joint distribution $G(\mathbf{y}, \mathbf{b})$ of \mathbf{y} and \mathbf{b} is replaced by the bootstrap distribution $\widetilde{G}(\mathbf{y}^*, \mathbf{b}^*)$ of \mathbf{y}^* and \mathbf{b}^* using (21). Then we define the bootstrap bias estimate b_c^* of the conditional log-likelihood as follows:

$$(26) \quad b_c^* = E_{\widetilde{G}(\mathbf{y}^*, \mathbf{b}^*)} \left[\log f_c(\mathbf{y}^*|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right] - E_{\widehat{G}_c(\mathbf{w})} \left[\log f_c(\mathbf{w}|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right].$$

So we define the bootstrap version of the conditional information criterion in LMM in the form of

$$(27) \quad \text{cEIC} = -2 \left(\log f_c(\mathbf{y}|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \widehat{\mathbf{K}}) - b_c^* \right).$$

As in the marginal case, to reduce the fluctuation in the bootstrap bias estimation of conditional log-likelihood, we also use the variance reduction method:

$$(28) \quad b_c^* \approx \frac{1}{B} \sum_{i=1}^B \{ D_{c1}(\mathbf{y}^*(i)) + D_{c3}(\mathbf{y}^*(i)) \},$$

where

$$\begin{aligned} D_{c1}(\mathbf{y}^*(i)) &= \log f_c(\mathbf{y}^*|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) - \log \left[f_c(\mathbf{y}^*|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right], \\ D_{c3}(\mathbf{y}^*(i)) &= E_{\widehat{G}_c} \log \left[f_c(\mathbf{w}|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \widehat{\mathbf{K}}) \right] - E_{\widehat{G}_c} \log \left[f_c(\mathbf{w}|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right]. \end{aligned}$$

3.3 Higher-order bias corrected information criteria

In this section, we focus on the second-step bootstrap bias corrected estimators of the log-likelihood and propose second-order bootstrap bias corrected information criteria. By now we have available four types of bias estimators, in which the bias estimators in mAIC and cAIC can be calculated analytically while the bias estimators in mEIC and cEIC can only be calculated numerically.

In mAIC, from (4) we know that the first-order bias estimate of the marginal log-likelihood is $p + 1$, so the second-order bootstrap bias estimate of the marginal log-likelihood in estimating the expected marginal log-likelihood is given by

$$(29) \quad \frac{1}{n} b_{2m}^* = E_{\widetilde{G}(\mathbf{y}^*)} \left[\log f_m(\mathbf{y}^*|\hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) - (p + 1) - E_{\widehat{G}(\mathbf{w})} \log f_m(\mathbf{w}|\hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right]$$

resulting the second-order bootstrap information criterion

$$(30) \quad \text{mAIC}_2 = n \log(2\pi\hat{\sigma}^2) + n + \log |\widehat{\mathbf{V}}| + 2 \left(p + 1 + \frac{1}{n} b_{2m}^* \right).$$

While in cAIC, from equation (9) we know that the first-order bias estimate of the conditional log-likelihood is given by $\rho + 1$, so the second-order bootstrap bias estimate of the conditional log-likelihood is given by

$$(31) \quad \frac{1}{n} b_{2c}^* = E_{\widetilde{G}(\mathbf{y}^*, \mathbf{b}^*)} \left[\log f_c(\mathbf{y}^*|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) - (\rho + 1) - E_{\widehat{G}_c(\mathbf{w})} \left[\log f_c(\mathbf{w}|\hat{\mathbf{b}}^*, \hat{\boldsymbol{\beta}}^*, \hat{\sigma}^{2*}, \widehat{\mathbf{K}}^*) \right] \right].$$

giving the second-order conditional bootstrap information criterion

$$(32) \quad \text{cAIC}_2 = -2 \log f_c(\mathbf{y}|\hat{\mathbf{b}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \widehat{\mathbf{K}}) + 2 \left(\rho + 1 + \frac{1}{n} b_{2c}^* \right).$$

Alternatively, as in defining mEIC and cEIC, we may first use the first-order bias correction of the log-likelihood numerically by bootstrap methods. Let b_m^* be defined as in (22). We estimate the second-order bias of the marginal log-likelihood as follows

$$(33) \quad \frac{1}{n}b_{2m}^{**} = E_{\tilde{G}(\mathbf{y}^*)} \left[\log f_m(\mathbf{y}^* | \hat{\beta}^*, \hat{\sigma}^{2*}, \hat{\mathbf{K}}^*) - b_m^* - E_{\hat{G}(\mathbf{w})} \log f_m(\mathbf{w} | \hat{\beta}^*, \hat{\sigma}^{2*}, \hat{\mathbf{K}}^*) \right],$$

giving the corresponding bootstrap information criterion as

$$(34) \quad \text{mEIC}_2 = -2 \left(\log f_m(\mathbf{y} | \hat{\beta}, \hat{\sigma}^2, \hat{\mathbf{K}}) - b_m^* - \frac{1}{n}b_{2m}^{**} \right).$$

Similarly to get second-order version of cEIC, we use the first-order bootstrap bias estimate b_c^* of the conditional log-likelihood defined by (26). The second-order bootstrap bias estimate of the conditional log-likelihood is then given by

$$(35) \quad \frac{1}{n}b_{2c}^{**} = E_{\tilde{G}(\mathbf{y}^*, \mathbf{b}^*)} \left[\log f_c(\mathbf{y}^* | \hat{\mathbf{b}}^*, \hat{\beta}^*, \hat{\sigma}^{2*}, \hat{\mathbf{K}}^*) - b_c^* - E_{\hat{G}_c(\mathbf{w})} \left[\log f_c(\mathbf{w} | \hat{\mathbf{b}}^*, \hat{\beta}^*, \hat{\sigma}^{2*}, \hat{\mathbf{K}}^*) \right] \right].$$

Thus we may define the second-order version of cEIC as follows

$$(36) \quad \text{cEIC}_2 = -2 \left(\log f_c(\mathbf{y} | \hat{\mathbf{b}}, \hat{\beta}, \hat{\sigma}^2, \hat{\mathbf{K}}) - b_c^* - \frac{1}{n}b_{2c}^{**} \right).$$

References

- [1] Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, **97**, 1-17.
- [2] Ishiguro, M., Sakamoto, Y., and Kitagawa, G. (1997). Bootstrapping log-likelihood and EIC, an extension of AIC. *Annals of the Institute of Statistical Mathematics*, **49**, 411-434.
- [3] Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika*, **83**, 875-890.
- [4] Konishi, S. and Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection-functional approach. *Journal of Statistical Planning and Inference*, **114**, 45-61.
- [5] Konishi, S., and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer: New York.
- [6] Kitagawa, G. and Konishi, S. (2010). Bias and variance reduction techniques for bootstrap information criteria. *Ann. Inst. Stat. Math.*, **62**, 209-234.
- [7] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963-974.
- [8] Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, **95**, 773-8.
- [9] Ngo, L. and Brand, R. (2002). Model selection in linear mixed effects models using SAS Proc Mixed. SUGI 22.
- [10] Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*, New York: Springer.
- [11] Srivastava M. S. and Kubokawa, T. (2010). Conditional information criteria for selecting variables in linear mixed models. *Journal of Multivariate Analysis*, **101**, 1970-1980.
- [12] Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351-70.

RÉSUMÉ (ABSTRACT)

In this paper we propose bootstrap information criteria for the linear mixed model. These information criteria are constructed using either the marginal log-likelihood or the conditional log-likelihood. Second-order bias-corrected information criteria are also considered.