An Assumption for the Development of Bootstrap Variants of the Akaike Information Criterion in Mixed Models

Junfeng Shang [1,*] and **Joseph E. Cavanaugh** [2,†]
[1]Bowling Green State University, USA
[2]The University of Iowa, USA

Abstract: This note provides a proof of a fundamental assumption in the verification of bootstrap AIC variants in mixed models. The assumption links the bootstrap data and the original sample data via the log-likelihood function, and is the key condition used in the validation of the criterion penalty terms. (See Assumption 3 of both Shibata, 1997, and Shang and Cavanaugh, 2007.) To state the assumption, let Y and Y^* represent the response vector and the corresponding bootstrap sample, respectively. Let θ represent the set of parameters for a candidate mixed model, and let $\hat{\theta}$ denote the corresponding maximum likelihood estimator based on maximizing the likelihood $L(\theta \mid Y)$. With E_* denoting the expectation with respect to the bootstrap distribution of Y^*, the assumption asserts that $E_* \log L(\hat{\theta} \mid Y^*) = \log L(\hat{\theta} \mid Y)$. We prove the assumption holds under parametric, semiparametric, and nonparametric bootstrapping.

Key words: AIC bootstrap variants, Bootstrap expectation, Log-likelihood function

---

[*] Corresponding author. Phone: (419) 372-7457. Fax: (419) 372-6092.
e-mail: jshang@bgnet.bgsu.edu. Department of Mathematics and Statistics, 450 Math Science Building, Bowling Green State University, Bowling Green, OH 43403.

[†] e-mail: joe-cavanaugh@uiowa.edu. Department of Biostatistics, The University of Iowa, IA 52242.

1 Introduction

Linear mixed models are utilized extensively in the modelling of correlated data. The selection of an appropriate mean and covariance structure is a fundamental challenge in the mixed modelling framework. Mixed model selection is facilitated by the application of a model selection criterion. The Akaike (1973, 1974) information criterion, AIC, is currently employed for mixed model selection in most situations. However, in settings where the sample size is small, AIC is likely to favor models of an inappropriately high dimension (e.g., Shang and Cavanaugh, 2007).

To improve the performance of AIC, the bootstrap can be used to develop AIC variants with refined penalty terms based on resampling. Such variants tend to outperform AIC in small-sample applications.

The idea of using the bootstrap to develop selection criteria was introduced by Efron (1983, 1986), and is extensively discussed by Efron and Tibshirani (1993, Chapter 17, pp. 237-253). Cavanaugh and Shumway (1997) proposed a bootstrap variant of AIC, AICb, for state-space model selection. Shibata (1997) has indicated the existence of five asymptotically equivalent bootstrap-corrected AIC variants, including a variant suggested by the work of Efron (AICb1) and the criterion of Cavanaugh and Shumway (AICb2). In the framework of the mixed model, AICb1 and AICb2 have been justified by Shang and Cavanaugh (2007).

In the justification of bootstrap AIC variants in mixed models, an assumption which links the bootstrap data and the original sample data via the log-likelihood function is crucial for the validation of the penalty terms. (See Assumption 3 of both Shibata, 1997, and Shang and Cavanaugh, 2007.) This paper provides a proof for this assumption in the mixed modelling framework, and presents a brief discussion of its implications. We also present the algorithms for parametric, semiparametric, and nonparametric bootstrapping.

2 Model and Notation

The repeated measures model, or the general linear mixed model, can be defined as

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i\,, \qquad i = 1, ..., m, \tag{2.1}$$

where y_i, β, b_i, and ε_i are all vectors, and X_i and Z_i are matrices.

Specifically, y_i denotes an $n_i \times 1$ vector of n_i responses observed on the ith subject; X_i is an $n_i \times (p+1)$ design matrix of full column rank; Z_i is an $n_i \times q$ design matrix; β is a $(p+1) \times 1$ fixed effects parameter vector; b_i is a $q \times 1$ random effects vector distributed as $N(0, \mathrm{D})$; and ε_i is an $n_i \times 1$ error vector distributed as $N(0, \sigma^2 R_i)$. D and R_i are $q \times q$ and $n_i \times n_i$ positive definite matrices, respectively, and σ^2 is a positive scalar. It is assumed that b_i and ε_i are distributed independently (and independently of one another) for $i = 1, ..., m$.

We regard case i as the response vector for the ith subject, y_i. Thus, the total number of cases is m, and the total number of observations is denoted by $N = \sum_{i=1}^{m} n_i$.

In the preceding model, the fixed effects parameters β need to be estimated, and the random effects b_i need to be predicted. Generally, D will consist of variance parameters which need to be estimated; R_i will be known for $i = 1, ..., m$, but σ^2 will need to be estimated.

A more succinct representation of model (2.1) can be obtained by combining all m subject-specific models into one overall model. This model will have the form

$$Y = X\beta + Zb + \varepsilon. \tag{2.2}$$

Here, Y denotes the $N \times 1$ response vector $({y_1}', \ldots, {y_m}')'$; X is an $N \times (p+1)$ design matrix of full column rank; Z is an $N \times mq$ block diagonal design matrix comprised of m blocks, where each block is an $n_i \times q$ matrix; β is the $(p+1) \times 1$ fixed effects parameter vector; b is the $mq \times 1$ random effects vector $({b_1}', \ldots, {b_m}')'$; and ε is the $N \times 1$ error vector $({\varepsilon_1}', \ldots, {\varepsilon_m}')'$.

We assume $b \sim N(0, D)$ and $\varepsilon \sim N(0, \sigma^2 R)$, with b and ε distributed independently. Here, R and D are positive definite block diagonal matrices; R is $N \times N$ and comprised of the m blocks $R_1, ..., R_m$, and D is $mq \times mq$ and comprised of m identical blocks, each of which is D.

Let θ denote the unknown parameter vector, consisting of the elements of the vector β, the matrix D, and the scalar σ^2. Let

$$V = ZDZ' + \sigma^2 R.$$

Note that V is the covariance matrix of Y and that V is positive definite.

To represent the likelihood corresponding to model (2.2), we use $L(\theta \mid Y)$. We use $\hat{\theta}$ to denote the maximum likelihood estimator (MLE) of θ based on maximizing $L(\theta \mid Y)$.

Let Y^* represent a bootstrap sample and let E_* represent the expectation with respect to the bootstrap distribution, i.e., with respect to the distribution of Y^*. Moreover, let $L(\theta \mid Y^*)$ represent the likelihood corresponding to model (2.2) under the bootstrap sample Y^*.

We use $\hat{\theta}^*$ to denote the bootstrap MLE of θ based on maximizing $L(\theta \mid Y^*)$. Accordingly, we use $\{\hat{\theta}^*(i),\, i = 1, \ldots, W\}$ to denote a set of W bootstrap replicates of $\hat{\theta}$.

The MLE vector $\hat{\theta}$ consists of MLEs based on the original sample Y: the vector $\hat{\beta}$, the matrix $\hat{\mathrm{D}}$, and the scalar $\hat{\sigma}^2$. Analogously, the bootstrap MLE vector $\hat{\theta}^*$ consists of the MLEs based on the bootstrap $\hat{\theta}^*$: the vector $\hat{\beta}^*$, the matrix $\hat{\mathrm{D}}^*$, and the scalar $\hat{\sigma}^{2*}$.

To generate a bootstrap sample, we resample on a case-by-case basis. The bootstrap sample size is taken to be the same as the size of the observed sample Y (i.e., m). The properties of the bootstrap when the bootstrap sample size is equal to the original sample size are discussed by Efron and Tibshirani (1993).

In the mixed modelling framework, we verify that

$$E_* \log L(\hat{\theta} \mid Y^*) = \log L(\hat{\theta} \mid Y) \tag{2.3}$$

holds under parametric, semiparametric, and nonparametric bootstrapping. As previously mentioned, this is the key condition used in the validation of the penalty terms for bootstrap AIC variants. Note that in practice, the bootstrap expectation $E_*(\cdot)$ is approximated by empirically averaging over a large collection of bootstrap replicates of the data.

Prior to the presentation of the proof of (2.3), we outline the algorithms of three bootstrapping procedures: parametric, semiparametric, and nonparametric.

3 Algorithms: Parametric, Semiparametric, and Nonparametric Bootstrapping

In this section, we will describe the algorithms of parametric, semiparametric, and nonparametric bootstrapping for generating the bootstrap samples and estimating the parameters. Based on these algorithms, assumption (2.3) can be verified.

In outlining the bootstrapping algorithms, we assume that the number of observations is the same for each case, i.e., $n_i = n$ for all i. We also assume that the $n \times q$ design matrix for the random effects for each case is the same for all i, i.e., $Z_i = \mathrm{Z}$, and that R_i is the same $n \times n$ identity matrix for all i, i.e., $R_i = \mathrm{I}$. In many practical applications, Z_i is a constant and R_i is the identity (e.g., repeated measures models with a compound symmetric covariance structure).

We remark that the EM-algorithm provides a convenient recourse for obtaining the MLEs. Other algorithms for maximizing the log likelihood may also be employed.

Parametric Bootstrapping

(1) Fit the mixed model (2.2) to the data to obtain the MLEs $\hat{\beta}$, $\hat{\mathrm{D}}$, and $\hat{\sigma}^2$.

(2) Generate the bootstrap sample case-by-case using the fitted model

$$y_i^* = X_i\hat{\beta} + \mathrm{Z}\hat{b}_i^* + \hat{\varepsilon}_i^*\,, \qquad i = 1, ..., m,$$

where $\hat{b}_i^*$ and $\hat{\varepsilon}_i^*$ are generated from $N(0, \hat{\mathrm{D}})$ and $N(0, \hat{\sigma}^2\mathrm{I})$ distributions, respectively.

(3) Fit the mixed model (2.2) to the bootstrap data, thereby obtaining the bootstrap MLEs $\hat{\beta}^*$, $\hat{\mathrm{D}}^*$, and $\hat{\sigma}^{2*}$.

(4) Repeat steps (2)-(3) W times.

Semiparametric Bootstrapping

The algorithm for semiparametric bootstrapping involves the residuals that adjust for only the fixed effects (i.e., the fitted mean structure). Once the MLEs are obtained, these residuals can be easily found via

$$\hat{\xi}_i = y_i - X_i\hat{\beta}, \qquad i = 1, \cdots, m. \tag{3.1}$$

The semiparametric bootstrap procedure for the mixed model (2.2) can be outlined as follows.

(1) As with step (1) for parametric bootstrapping, obtain the MLE's $\hat{\beta}$, $\hat{\mathrm{D}}$, and $\hat{\sigma}^2$.

(2) Obtain the residuals $\hat{\xi}_1, \ldots, \hat{\xi}_m$ using (3.1).

(3) Draw a sample of size m with replacement from these residuals. Denote the bootstrap residuals by $\hat{\xi}_1^*, \ldots, \hat{\xi}_m^*$.

(4) Construct the bootstrap data set using the fitted model

$$y_i^* = X_i\hat{\beta} + \hat{\xi}_i^*\,, \qquad i = 1, \ldots, m.$$

(5) Fit the mixed model (2.2) to the bootstrap data, thereby obtaining the bootstrap MLEs $\hat{\beta}^*$, $\hat{\mathrm{D}}^*$, and $\hat{\sigma}^{2*}$.

(6) Repeat steps (3)-(5) W times.

Morris (2002) presented a slightly different semiparametric bootstrapping procedure for mixed models using the best linear unbiased predictors (BLUPs) of the random effects. Morris' procedure differs than the one outlined here in the residual formulation step (2), the resampling step (3), and the bootstrap data construction step (4). In Morris' algorithm, the BLUPs are utilized to recover the random effects. In step (2), residuals that approximate the error terms are evaluated by subtracting the BLUPs from the mean-adjusted residuals $\hat{\xi}_i^*$ in (3.1) on a case-by-case basis. In step (3), these error-term residuals are then resampled; the BLUPs are also resampled separately. In step (4), the bootstrap cases y_i^* are then generated by combining the mean estimates $X_i\hat{\beta}$ with the resampled BLUPs and the resampled error-term residuals.

Nonparametric Bootstrapping

(1) Match y_i and X_i to form the pairs of data structures $(y_i, X_i), i = 1, \cdots, m$. Draw a sample of size m with replacement from the m pairs. Denote the pairs in the bootstrap sample $(y_1^*, X_1^*), \cdots, (y_m^*, X_m^*)$.

(2) Fit the mixed model (2.2) to the bootstrap data, thereby obtaining the bootstrap MLEs $\hat{\beta}^*$, $\hat{\mathrm{D}}^*$, and $\hat{\sigma}^{2*}$.

(3) Repeat steps (1)-(2) W times.

4 Proof of the Assumption

We now establish (2.3) in the context of the linear mixed model (2.1), (2.2). We will prove that this assumption holds under parametric, semiparametric, and nonparametric bootstrapping.

Parametric Bootstrapping

With parametric bootstrapping, a bootstrap sample Y^* is produced via the formula

$$Y^* = X\hat{\beta} + \hat{\xi}^*,$$

where $\hat{\beta}$ is the MLE of β, and the bootstrap residuals $\hat{\xi}^*$ are generated from a normal distribution with mean vector 0 and covariance matrix $\hat{V} = Z\hat{D}Z' + \hat{\sigma}^2 R$. Note that $\hat{V}$ is the MLE of V under model (2.2).

Neglecting the constant in $\log L(\hat{\theta} \mid Y^*)$, we have

$$\begin{aligned}
E_* \log L(\hat{\theta} \mid Y^*) &= E_*[-\frac{1}{2}\log|\hat{V}| - \frac{1}{2}(Y^* - X\hat{\beta})'\hat{V}^{-1}(Y^* - X\hat{\beta})] \\
&= -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}E_*\|\hat{V}^{-\frac{1}{2}}(Y^* - X\hat{\beta})\|^2 \\
&= -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}E_*\|\hat{V}^{-\frac{1}{2}}\hat{\xi}^*\|^2 \\
&= -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}\mathrm{tr}(\hat{V}^{-1}\hat{V}) \\
&= -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}N.
\end{aligned}$$

Recall that N denotes the total number of observations in the vector Y.

Similarly, we have

$$\log L(\hat{\theta} \mid Y) = -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}(Y - X\hat{\beta})'\hat{V}^{-1}(Y - X\hat{\beta}).$$

From Christensen (1996, pp. 271-272), for the MLEs $\hat{V}$ and $\hat{\beta}$, we have

$$(Y - X\hat{\beta})'\hat{V}^{-1}(Y - X\hat{\beta}) = N.$$

Thus,

$$\log L(\hat{\theta} \mid Y) = -\frac{1}{2}\log|\hat{V}| - \frac{1}{2}N.$$

Therefore, the result holds under parametric bootstrapping.

Semiparametric Bootstrapping

Neglecting the constant in $\log L(\hat\theta \mid Y^*)$, we again have

$$\begin{aligned} E_* \log L(\hat\theta \mid Y^*) &= E_*[-\frac{1}{2}\log|\hat V| - \frac{1}{2}(Y^* - X\hat\beta)'\hat V^{-1}(Y^* - X\hat\beta)] \\ &= -\frac{1}{2}\log|\hat V| - \frac{1}{2}E_*\|\hat V^{-\frac{1}{2}}(Y^* - X\hat\beta)\|^2. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \log L(\hat\theta \mid Y) &= -\frac{1}{2}\log|\hat V| - \frac{1}{2}(Y - X\hat\beta)'\hat V^{-1}(Y - X\hat\beta) \\ &= -\frac{1}{2}\log|\hat V| - \frac{1}{2}\|\hat V^{-\frac{1}{2}}(Y - X\hat\beta)\|^2. \end{aligned}$$

Therefore, we need only show that

$$E_*\|\hat V^{-\frac{1}{2}}(Y^* - X\hat\beta)\|^2 = \|\hat V^{-\frac{1}{2}}(Y - X\hat\beta)\|^2. \tag{4.1}$$

With reference to (3.1), we have

$$y_i - X_i\hat\beta = \hat\xi_i, \qquad i = 1, \ldots, m,$$

where $\hat\xi_1, \ldots, \hat\xi_m$ are the case-specific residuals. Let $\hat\xi$ denote the $N \times 1$ vector $(\hat\xi_1', \ldots, \hat\xi_m')'$.

With semiparametric bootstrapping, a bootstrap sample Y^* is generated via the formula

$$Y^* = X\hat\beta + \hat\xi^*,$$

where $\hat\beta$ is the MLE of β, and the bootstrap residuals $\hat\xi^*$ are generated based on the empirical distribution of the residuals $\hat\xi = Y - X\hat\beta$. Note that $\hat\xi^*$ denotes the $N \times 1$ vector $(\hat\xi_1^{*\prime}, \ldots, \hat\xi_m^{*\prime})'$.

Let V^* denote the covariance matrix of $\hat\xi^*$. Under semiparametric bootstrapping, the distribution of $\hat\xi^*$ is the same as the empirical distribution of $\hat\xi$. With $\bar{\hat\xi} = \sum_{i=1}^m \hat\xi_i/m$, we therefore see that V^* is a positive definite block diagonal matrix comprised of m identical blocks, where each block has the form $\frac{1}{m}\sum_{i=1}^m(\hat\xi_i - \bar{\hat\xi})(\hat\xi_i - \bar{\hat\xi})'$. Let $\bar\xi$ denote the $N \times 1$ vector $(\bar{\hat\xi}', \ldots, \bar{\hat\xi}')'$. Then, the mean of $\hat\xi^*$ is given by $\bar\xi$.

Now, to show equation (4.1) holds, we will utilize the relation

$$(\hat{\xi}-\bar{\xi})'\hat{V}^{-1}(\hat{\xi}-\bar{\xi})+\bar{\xi}'\hat{V}^{-1}\bar{\xi}=\hat{\xi}'\hat{V}^{-1}\hat{\xi}.$$

This relation will be established in the following Lemma.

Starting with the left-hand side of (4.1), we obtain

$$\begin{aligned} E_*\|\hat{V}^{-\frac{1}{2}}(Y^*-X\hat{\beta})\|^2 &= E_*\|\hat{V}^{-\frac{1}{2}}\hat{\xi}^*\|^2 \\ &= \operatorname{tr}(\hat{V}^{-1}V^*)+[E_*(\hat{\xi}^*)]'\hat{V}^{-1}E_*(\hat{\xi}^*) \\ &= (\hat{\xi}-\bar{\xi})'\hat{V}^{-1}(\hat{\xi}-\bar{\xi})+\bar{\xi}'\hat{V}^{-1}\bar{\xi} \\ &= \hat{\xi}'\hat{V}^{-1}\hat{\xi} \\ &= \|\hat{V}^{-\frac{1}{2}}\hat{\xi}\|^2 \\ &= \|\hat{V}^{-\frac{1}{2}}(Y-X\hat{\beta})\|^2. \end{aligned}$$

Thus, (4.1) is established for semiparametric bootstrapping.

Lemma 1

$$(\hat{\xi}-\bar{\xi})'\hat{V}^{-1}(\hat{\xi}-\bar{\xi})+\bar{\xi}'\hat{V}^{-1}\bar{\xi}=\hat{\xi}'\hat{V}^{-1}\hat{\xi}.$$

Proof: First, we have

$$\begin{aligned} &(\hat{\xi}-\bar{\xi})'\hat{V}^{-1}(\hat{\xi}-\bar{\xi})+\bar{\xi}'\hat{V}^{-1}\bar{\xi} \\ =\;& \hat{\xi}'\hat{V}^{-1}\hat{\xi}-\hat{\xi}'\hat{V}^{-1}\bar{\xi}-\bar{\xi}'\hat{V}^{-1}\hat{\xi}+\bar{\xi}'\hat{V}^{-1}\bar{\xi}+\bar{\xi}'\hat{V}^{-1}\bar{\xi} \\ =\;& \hat{\xi}'\hat{V}^{-1}\hat{\xi}-2\hat{\xi}'\hat{V}^{-1}\bar{\xi}+2\bar{\xi}'\hat{V}^{-1}\bar{\xi}. \end{aligned}$$

Thus, it suffices to prove that

$$\hat{\xi}'\hat{V}^{-1}\bar{\xi}=\bar{\xi}'\hat{V}^{-1}\bar{\xi}. \tag{4.2}$$

Since we assume that the covariance matrix for each case is the same and that the cases are independent, we have that $\hat{V}^{-1}$ is a block diagonal matrix comprised of m identical blocks, where each block has the form $\hat{\mathrm{V}}^{-1}$. (Specifically, $\hat{\mathrm{V}}^{-1} = (\mathrm{Z}'\hat{\mathrm{D}}\mathrm{Z} + \hat{\sigma}^2\mathrm{I})^{-1}$.) Then, the left-hand side of (4.2) can be expressed as

$$\hat{\xi}'\hat{V}^{-1}\bar{\xi} = \sum_{i=1}^{m} \hat{\xi}_i'\hat{\mathrm{V}}^{-1}\bar{\hat{\xi}} = \mathrm{tr}(\hat{\mathrm{V}}^{-1}\bar{\hat{\xi}}\sum_{i=1}^{m}\hat{\xi}_i') = \mathrm{tr}(m\hat{\mathrm{V}}^{-1}\bar{\hat{\xi}}\bar{\hat{\xi}}').$$

Similarly, the right-hand side of (4.2) can be expressed as

$$\bar{\xi}'\hat{V}^{-1}\bar{\xi} = \sum_{i=1}^{m} \bar{\hat{\xi}}'\hat{\mathrm{V}}^{-1}\bar{\hat{\xi}} = \mathrm{tr}(m\hat{\mathrm{V}}^{-1}\bar{\hat{\xi}}\bar{\hat{\xi}}').$$

Therefore, we can see that equation (4.2) holds and the proof of the lemma is complete.

For nonparametric bootstrapping, we will prove that

$$E_* \log L(\theta \mid Y^*) = \log L(\theta \mid Y) \tag{4.3}$$

holds for any value of θ. The result (2.3) then holds as a special case, where $\theta = \hat{\theta}$.

Nonparametric Bootstrapping

In nonparametric bootstrapping, we match y_i and X_i to form the pairs of data structures $(y_i, X_i), i = 1, \cdots, m$, and then draw a sample of size m with replacement from the m pairs. We denote the pairs in the bootstrap sample $(y_1^*, X_1^*), \cdots, (y_m^*, X_m^*)$. Therefore, a bootstrap sample is comprised of a set of m pairs (y_i^*, X_i^*), $i = 1, \ldots, m$, randomly drawn from the pairs (y_i, X_i), $i = 1, \ldots, m$. As before, Y^* denotes the $N \times 1$ vector $(y_1^{*\prime}, \ldots, {y_m^*}')'$; we let X^* represent an $N \times (p+1)$ design matrix defined as $X^* = [X_1^{*\prime} \cdots {X_m^*}']'$.

With reference to the case-by-case model (2.1), we have

$$y_i - X_i\beta \;=\; \xi_i, \qquad i = 1, \ldots, m,$$

where $\xi_i = Z_i b_i + \varepsilon_i$. With reference to the composite model (2.2), we have

$$\begin{aligned} Y - X\beta &= \xi, \text{ and} \\ Y^* - X^*\beta &= \xi^*, \end{aligned} \tag{4.4}$$

where $\xi = Zb + \varepsilon$. Note that ξ denotes the $N \times 1$ vector $(\xi_1', \ldots, \xi_m')'$.

Let V^* denote the covariance matrix of ξ^*. As implied by (4.4), under nonparametric bootstrapping, for any θ, the bootstrap distribution of ξ^* is the same as the empirical distribution of ξ. Without the loss of generality, we assume the mean of the model error vectors $\xi_1, \ldots, \xi_m$ is zero. The general proof that doesn't impose this zero-mean assumption is based on arguments that are similar to those used in the proof for semiparametric bootstrapping. For brevity of exposition, we employ this simplifying assumption. We therefore have that V^* is a positive definite block diagonal matrix comprised of m identical blocks, where each block has the form $\frac{1}{m}\sum_{i=1}^{m} \xi_i \xi_i'$.

To establish the result (4.3) for nonparametric bootstrapping, we use the following definition of the log likelihood in place of $\log L(\theta \mid Y^*)$:

$$\begin{aligned}\log L(\theta \mid (Y^*, X^*)) &= -\frac{1}{2}\log|V| - \frac{1}{2}(Y^* - X^*\beta)'V^{-1}(Y^* - X^*\beta)\\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\|V^{-\frac{1}{2}}(Y^* - X^*\beta)\|^2.\end{aligned}$$

The preceding is a well-known definition of the log likelihood suitable for nonparametric bootstrapping (Efron and Tibshirani, 1993). However, for consistency of notation, we use $\log L(\theta \mid Y^*)$ in place of $\log L(\theta \mid (Y^*, X^*))$, even when nonparametric bootstrapping is employed.

For the original log likelihood, we have

$$\begin{aligned}\log L(\theta \mid Y) &= -\frac{1}{2}\log|V| - \frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta)\\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\|V^{-\frac{1}{2}}(Y - X\beta)\|^2.\end{aligned}$$

Therefore, we need only to show that

$$E_*\|V^{-\frac{1}{2}}(Y^* - X^*\beta)\|^2 = \|V^{-\frac{1}{2}}(Y - X\beta)\|^2. \tag{4.5}$$

Using (4.4) and the form of V^*, we obtain

$$\begin{aligned} E_* \| V^{-\frac{1}{2}} (Y^* - X^* \beta) \|^2 &= E_* \| V^{-\frac{1}{2}} \xi^* \|^2 \\ &= \mathrm{tr}(V^{-1} V^*) \\ &= \xi' V^{-1} \xi \\ &= \| V^{-\frac{1}{2}} \xi \|^2 \\ &= \| V^{-\frac{1}{2}} (Y - X\beta) \|^2. \end{aligned}$$

Thus, (4.5) is established for nonparametric bootstrapping.

We have therefore demonstrated that the relation (2.3) holds under parametric, semiparametric, and nonparametric bootstrapping for the general linear mixed model defined in (2.1) and (2.2), under the common assumption of constant n_i, constant Z_i, and identity R_i.

5 Discussion

The relation $E_* \log L(\hat{\theta} \mid Y^*) = \log L(\hat{\theta} \mid Y)$ bridges the bootstrap sample and the original sample by the log likelihood at the MLE parameter point. This assumption facilitates the justification of bootstrap-based penalty terms of AIC variants.

In the mixed modelling framework, traditional AIC may be expressed as

$$\mathrm{AIC} = -2 \, \log L(\hat{\theta} \mid Y) + 2k,$$

where k denotes the dimension of $\hat{\theta}$, which includes both the regression parameters associated with the fixed effects and the variance/covariance parameters associated with the random effects. The "goodness of fit" term, $-2 \log L(\hat{\theta} \mid Y)$, serves as a biased estimator of the expected Kullback-Leibler discrepancy between the generating model and a fitted candidate model. The penalty term, $2k$, estimates the bias adjustment, which is represented by the difference between the expected discrepancy and the expected "goodness of fit" term.

AIC serves as an approximately unbiased estimator of the expected Kullback-Leibler discrepancy in settings where the sample size is large and k is relatively small. In settings where the sample size is small and k is relatively large, the penalty term $2k$ is often much smaller than the bias adjustment, making AIC substantially negatively biased as an estimator of the expected discrepancy. As a result, the criterion may favor overspecified models even when the expected discrepancy between these models and the generating model is quite large.

With assumption (2.3), one can develop and justify bootstrap-based AIC variants to adjust for this weakness in the penalty term of AIC. As previously mentioned, two such variants were suggested by the work of Efron (1983, 1986) and Cavanaugh and Shumway (1997). In the present setting, these variants can be respectively expressed as follows:

$$\begin{aligned} \text{AICb1} &= -2\,\log L(\hat{\theta} \mid Y) + \frac{1}{W}\sum_{i=1}^{W} -2\log\frac{L(\hat{\theta}^{*}(i) \mid Y)}{L(\hat{\theta}^{*}(i) \mid Y^{*}(i))}, \quad \text{and} \\ \text{AICb2} &= -2\,\log L(\hat{\theta} \mid Y) + 2\left\{\frac{1}{W}\sum_{i=1}^{W} -2\log\frac{L(\hat{\theta}^{*}(i) \mid Y)}{L(\hat{\theta} \mid Y)}\right\}. \end{aligned}$$

Shibata (1997) proved assumption (2.3) in the context of linear regression and considered AICb1 and AICb2 in addition to three additional bootstrap-corrected AIC variants. In the mixed modelling framework, AICb1 and AICb2 are asymptotically equivalent, and provide asymptotically unbiased estimators of the expected Kullback-Leibler discrepancy between the generating model and a fitted candidate model. Simulation results demonstrate that these two variants outperform AIC both as estimators of the expected discrepancy and as selection criteria. See Shang and Cavanaugh (2007) for details.

In the justification of these two bootstrap variants of AIC in mixed models, the assumption established here not only connects bootstrap and conventional log likelihoods, but also connects the corresponding score functions and information matrices, which therefore leads to asymptotic equivalencies based on second-order Taylor series expansions. Relying on such equivalencies, one can establish that AICb1 and AICb2 are asymptotically unbiased estimators of the expected Kullback-Leibler discrepancy, and that the criteria are asymptotically equivalent. Therefore, the assumption established in this paper is key to the justification of AICb1 and AICb2 in the mixed modelling framework, as well as the justification of other bootstrap-based AIC variants of analogous form.

Acknowledgement

The authors wish to express their appreciation to an anonymous referee for valuable suggestions that served to improve the original version of this manuscript.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki ed., 2nd International Symposium on Information Theory, pp. 267–281. Akademia Kiado, Budapest.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19, 716–723.

Cavanaugh, J.E., Shumway, R.H., 1997. A bootstrap variant of AIC for state-space model selection. Statistica Sinica 7, 473–496.

Christensen, R., 1996. Plane Answers to Complex Questions: The Theory of Linear Models (Second Edition). Springer, New York.

Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. Journal of the American Statistical Association 78, 316–331.

Efron, B., 1986. How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association 81, 461–470.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Morris, J.S., 2002. The BLUPs are not "best" when it comes to bootstrapping. Statistics & Probability Letters 56, 425-430.

Shang, J., Cavanaugh, J.E., 2007. Bootstrap variants of the Akaike information Criterion for Mixed Model Selection. Computational Statistics and Data Analysis. In press.

Shibata, R., 1997. Bootstrap estimate of Kullback-Leibler information for model selection. Statistica Sinica 7, 375–394.