Journal of Multivariate Analysis 116 (2013) 245–262

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Information based model selection criteria for generalized linear mixed models with unknown variance component parameters

Dalei Yu [a], Xinyu Zhang [b], Kelvin K.W. Yau [c,*]

[a] Statistics and Mathematics College, Yunnan University of Finance and Economics, Kunming 650221, China
[b] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 7 March 2012
Available online 11 January 2013

AMS subject classifications:
62J12
62F12
62F40

Keywords:
Conditional Akaike information
Model selection
Parametric bootstrap
Variance component

ABSTRACT

This paper derives the corrected conditional Akaike information criteria for generalized linear mixed models by analytic approximation and parametric bootstrap. The sampling variation of both fixed effects and variance component parameter estimators are accommodated in the bias correction term. Simulation shows that the proposed corrected criteria provide good approximation to the true conditional Akaike information and demonstrates promising model selection results. The use of the criteria is demonstrated in the analysis of the chronic asthmatic patients' data.

1. Introduction

The generalized linear mixed model (GLMM) is a powerful tool for the analysis of non-normal and non-independent data. When the research interest is on cluster effects in addition to population-level parameters, the traditional Akaike information criterion is not appropriate because the prediction of random effects is involved.

Motivated by this, Vaida and Blanchard [24] defined conditional Akaike information (cAI), and the corresponding conditional Akaike information criterion (cAIC) for linear mixed models (LMMs) was derived based on the best linear unbiased predictor (BLUP), where the variance component parameters of random effects were assumed to be known. Within the framework of cAI, Srivastava and Kubokawa [23] considered the problem of selecting the variables of the fixed effects in LMMs, and also assumed the known variance component parameters. When the variance component parameters are unknown, Vaida and Blanchard [24] and Srivastava and Kubokawa [23] suggested using the plug-in estimator for the unknown variance component parameters, the resulting criterion was termed as conventional cAIC by Greven and Kneib [9]. Moreover, Liang et al. [18] proposed a corrected version that accounts for the sampling variation of the variance component parameter estimators and showed the advantage of their corrected version over the conventional cAIC by simulation. In addition, Kubokawa [15] extended Liang's corrected cAIC to the situation when all the parameters are unknown. In comparison with the corrected cAIC, Greven and Kneib [9] proved that in LMM settings, ignoring the uncertainty in estimation of the variance component parameters induced a specific bias on the conventional cAIC and advocated the use of the corrected cAIC in practice.

* Corresponding author.
E-mail addresses: yudalei@126.com (D. Yu), xinyu@amss.ac.cn (X. Zhang), mskyau@cityu.edu.hk (K.K.W. Yau).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.12.005

Recently, Donohue et al. [7] extended cAIC to GLMMs not only because of their focus on clusters but also due to the fact that the marginal distribution of response usually has no explicit form. In Donohue et al. [7], the estimated variance component parameters were directly plugged into the expression criterion and the resulting cAIC can be viewed as an extension of conventional cAIC of LMM. In this paper, accommodating the sampling variation of variance component parameter estimators in the bias correction term, we shall develop two corrected forms of cAIC under the GLMM setting of Donohue et al. [7] by analytic approximation and bootstrap, respectively. We prove that under some regularity conditions, the difference between the conventional and corrected versions of cAIC is asymptotically ignorable. However, when the variance of random effects is very small, such difference in the finite sample situation becomes apparent. Simulation studies also show that our proposed corrected cAICs perform well in estimation of cAI and in model selection.

The rest of the paper is arranged as follows. Section 2 describes the model framework and notations. Section 3 derives the corrected cAICs for GLMMs where the corresponding bias correction term is obtained by analytic approximation and parametric bootstrap. The difference between corrected cAIC (analytic approximation) and the conventional cAIC is explored. Section 4 presents a simulation study on the behavior of the proposed criteria and demonstrates the method with an application to a set of chronic asthmatic patients' data. Section 5 concludes. Regularity conditions and technical proofs are presented in the Appendix.

2. Model framework and notations

Consider the data consisting of outcomes from m clusters, with n_i observations for the ith cluster ($i = 1, \ldots, m$). In the ith cluster, conditional on the cluster-specific $d \times 1$ vector of random effects b_i, the outcomes y_{ij} are independent and assumed to follow a GLMM with mean

$$E(y_{ij}|\beta, b_i) = g^{-1}(\eta_{ij}) = g^{-1}(\beta^{\mathrm{T}}x_{ij} + b_i^{\mathrm{T}}z_{ij}), \tag{1}$$

where β is the $p \times 1$ fixed effects vector, x_{ij} and z_{ij} are the vectors in the design matrix corresponding to fixed and random effects, respectively, and $g(\cdot)$ is the link function. Let $N = n_1 + \cdots + n_m, y_{i\cdot} = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}, y = \mathrm{stack}(y_{1\cdot}, \ldots, y_{m\cdot})$ (the stack function stacks matrices on the top of each other), $b = \mathrm{stack}(b_1, \ldots, b_m)$, $X = \mathrm{stack}(X_1, \ldots, X_m)$, $Z = \mathrm{diag}(Z_1, \ldots, Z_m)$, and $U = (X, Z)$. $\eta = (\eta_{11}, \ldots, \eta_{1n_1}, \ldots, \eta_{m1}, \ldots, \eta_{mn_m})^{\mathrm{T}}$.

Assume that $b_1, \ldots, b_m$ are independently distributed with $Eb_i = 0_{d\times 1}$ and $b_i \sim \kappa(b_i|\varphi)$, where φ is the υ-dimensional variance component parameter vector. Denote $\xi = \mathrm{stack}(\beta, \varphi)$ as the vector of unknown parameters and $l_i(y_{i\cdot} \mid \beta, b_i)$ as the log-likelihood function of the ith cluster conditional on random effects vector b_i, then the joint log-likelihood is

$$l_{\mathrm{joint}}(y, b \mid \xi) = \sum_{i=1}^{m} l_{\mathrm{joint},i}(y_{i\cdot}, b_i \mid \xi) = \sum_{i=1}^{m} \{l_i(y_{i\cdot} \mid \beta, b_i) + \log\kappa(b_i \mid \varphi)\}. \tag{2}$$

Let $q = d \times m, \theta_m = \mathrm{stack}(\beta, b_1, \ldots, b_m), l(y \mid \theta_m) = \sum_{i=1}^{m} l_i(y_{i\cdot} \mid \beta, b_i)$, and $l_\kappa(b \mid \varphi) = \sum_{i=1}^{m}\log\kappa(b_i \mid \varphi)$. When the model is correctly specified, we have

$$E_{y|b}\{\partial l(y \mid \theta_m)/\partial\theta_m\} = 0_{(p+q)\times 1}, \quad E_b\{\partial\log\kappa(b_i \mid \varphi)/\partial b_i\} = 0_{d\times 1}. \tag{3}$$

Denote $\tilde{b}_i(\xi)$ as the solution of $\partial l_{\mathrm{joint},i}(y_{i\cdot}, b_i \mid \xi)/\partial b_i = 0_{d\times 1}$ and $\tilde{\beta}(\varphi)$ as the solution of $\partial l_{\mathrm{joint}}(y, b \mid \xi)/\partial\beta|_{b_i=\tilde{b}_i(\xi)} = 0_{p\times 1}$. Let $\hat{\varphi}$ be a consistent estimator of φ, $\hat{\xi} = \mathrm{stack}(\hat{\beta}, \hat{\varphi})$, $\hat{\beta} = \tilde{\beta}(\hat{\varphi})$, and $\hat{b}_i = \tilde{b}_i(\hat{\xi})$. Then the estimator/predictor vector

$$\hat{\theta}_{nm} = \mathrm{stack}\left(\hat{\beta}, \hat{b}_1, \ldots, \hat{b}_m\right)$$

satisfies

$$\left.\frac{\partial l_{\mathrm{joint}}(y, b \mid \xi)}{\partial\theta_m}\right|_{\theta_m=\hat{\theta}_{nm}, \varphi=\hat{\varphi}} = 0_{(p+q)\times 1}. \tag{4}$$

Lee and Nelder [16] termed $\hat{\theta}_{nm}$ as the maximum hierarchical likelihood estimator (MHLE), and considered it as a generalization of the BLUP type estimator in non-normal settings. Moreover, it can also be verified that $\hat{\theta}_{nm}$ is the maximum posterior estimator [12] of θ_m. This estimator was widely discussed in the literature, such as Breslow and Clayton [5], McGilchrist [20] and Yau and Kuk [26], and its asymptotic property was studied in Lee and Nelder [16], Jiang et al. [12] and Nie [21].

Let $\hat{\nabla}'_\theta(y \mid \theta_m) = -\partial l(y \mid \theta_m)/\partial\theta_m$, $\hat{\nabla}''_{\theta\theta}(y \mid \theta_m) = -\partial^2 l(y \mid \theta)/\partial\theta_m\partial\theta_m^{\mathrm{T}}$, and $\nabla''_{\theta\theta}(\xi) = E_{y,b}\left\{\hat{\nabla}''_{\theta\theta}(y \mid \theta_m)\right\}$. Denote $\hat{\nabla}''_{\beta\beta}(y \mid \theta_m)$, $\hat{\nabla}''_{\beta b_i}(y \mid \theta_m)$ and $\hat{\nabla}''_{b_ib_i}(y \mid \theta_m)$ as the corresponding matrix partitions with respect to fixed/random effects vectors in $\hat{\nabla}''_{\theta\theta}(y \mid \theta_m)$. Similarly, $\nabla''_{\beta\beta}(\xi)$, $\nabla''_{\beta b_i}(\xi)$ and $\nabla''_{b_ib_i}(\xi)$ are sub-matrices of $\nabla''_{\theta\theta}(\xi)$. Let $\hat{\nabla}''_{\beta b}(y \mid \theta_m) = \left\{\hat{\nabla}''_{\beta b_1}(y \mid \theta_m), \ldots, \hat{\nabla}''_{\beta b_m}(y \mid \theta_m)\right\}$ and $\hat{\nabla}''_{bb}(y \mid \theta_m) = \mathrm{diag}\left\{\hat{\nabla}''_{b_1b_1}(y \mid \theta_m), \ldots, \hat{\nabla}''_{b_mb_m}(y \mid \theta_m)\right\}$.

Moreover, let $\hat{\Delta}'_\theta(y \mid \theta_m, \varphi) = -\partial l_{\mathrm{joint}}(y, b \mid \xi)/\partial\theta_m$, $\hat{\Delta}''_{\theta\theta}(y \mid \theta_m, \varphi) = -\partial^2 l_{\mathrm{joint}}(y, b \mid \xi)/\partial\theta_m\partial\theta_m^{\mathrm{T}}$, and $\hat{\Delta}''_{\beta\beta}(y \mid \theta_m, \varphi)$, $\hat{\Delta}''_{\beta b_i}(y \mid \theta_m, \varphi)$ and $\hat{\Delta}''_{b_ib_i}(y \mid \theta_m, \varphi)$ be the corresponding matrix blocks with respect to the fixed/random effects vectors of