Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Standard Error of the Method of Simulated Moment Estimator for Generalized Linear Mixed Models

Yan Lu

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA


To cite this article: Yan Lu (2013): Standard Error of the Method of Simulated Moment Estimator for Generalized Linear Mixed Models, Communications in Statistics - Simulation and Computation, 42:1, 1-7
To link to this article: http://dx.doi.org/10.1080/03610918.2011.625485

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Standard Error of the Method of Simulated Moment Estimator for Generalized Linear Mixed Models

YAN LU

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA

This article considers standard error estimation of the method of simulated moment estimator for generalized linear mixed models. In literature, parametric bootstrap is used to estimate the covariance matrix, in which we use the estimator to generate simulated moments. To avoid the bias introduced by estimating the parameter and to deal with the correlated observations, we propose a two-stage block nonparametric bootstrap to estimate the standard errors. It is shown from simulation study that the proposed method performs well.

Keywords Bootstrap; Generalized linear mixed model; Simulated moments; Simulations; Standard errors.

Mathematics Subject Classification 62-04; 62G05; 62G09.

1. Introduction

Generalized linear mixed models (GLMM) are extensions of generalized linear models (McCullagh and Nelder, 1989) by the inclusion of random effects in the predictor. An issue regarding GLMM is computational difficulty of solving the maximum likelihood estimator (MLE) because of the high-dimensional integrals in the likelihood function.

Penalized quasi-likelihood (PQL) estimation (Breslow and Clayton, 1993) and bias corrected PQL estimation (Lin and Breslow, 1996) are popular in solving the MLEs of GLMM. Jiang (1998) showed that the estimators from the above two methods are both inconsistent. Method of moments (MOM) is another way used to estimate the parameters for GLMM. An extension of MOM is called method of simulated moments (MSM). MSM approximates the moments by Monte Carlo simulation when direct computation of the moments are not possible. Jiang (2009) showed that MSM estimator of GLMM is consistent; the precision and efficiency of the MSM estimator are competitive to PQL type estimator while the computation is relatively simple.
The interest of this study is to estimate standard error of MSM estimator for GLMM. Jiang (2009) suggested a parametric bootstrap for estimating the covariance matrix. A problem of parametric bootstrap is that the estimators are used to generate the simulated moments instead of the true parameters, in which bias is introduced by estimating the parameters. To avoid the bias and deal with the correlated observations, we propose a two-stage block nonparametric bootstrap to estimate the standard errors.

The present article is organized into five sections. In Sec. 2, we discuss the background of this study including GLMM and MSM. In Sec. 3, we propose a two-stage block nonparametric bootstrap method to estimate the standard errors. Simulation study is reported in Sec. 4. Finally, we give a summary in Sec. 5.

2. Background

2.1. Generalized Linear Mixed Models (GLMM)

GLMM is an extension of the general linear mixed models and generalized linear models. The responses in GLMM are correlated and categorical. Given a vector of random effects \( \mathbf{z} = (z_1, z_2, \ldots, z_n) \), the observations \( y_1, y_2, \ldots, y_n \) are conditionally independent such that \( y_i \sim N(x_i^\prime \beta + z_i^\prime \alpha, \tau^2) \), where \( y_i \) is a vector of observations related to random effect \( z_i \), \( x_i \) and \( z_i \) are known vectors, \( \beta \) is an unknown \( p \times 1 \) vector of regression parameters, and \( \tau^2 \) is an unknown variance. Furthermore, suppose that \( \mathbf{z} \) is multivariate normal with mean \( \mathbf{0} \) and covariance matrix \( \mathbf{G} \). Let \( \mathbf{X} \) and \( \mathbf{Z} \) be the matrices whose \( i \)th rows are \( x_i \) and \( z_i \), respectively. \( y \) can be expressed as \( y = X\beta + Z\alpha + \epsilon \), where \( \epsilon \) is multivariate normal with mean \( \mathbf{0} \) and covariance matrix \( \tau^2 \mathbf{I} \). Through this article, the responses \( y_i \) are correlated in nonoverlapping blocks (or strata, clusters or groups) related to random effect \( z_i \). We assume that the population blocks are all essentially infinite. The distribution of \( y_i \) does not depend on \( n_i \), the size of the sample taken from the \( i \)th block.

To define a GLMM, suppose that, given a vector of random effects \( \mathbf{z} \), the responses \( y_1, y_2, \ldots, y_n \) are conditionally independent such that the conditional distribution of \( y_i \) given \( \mathbf{z} \) is a member of the exponential family with the following probability density function:

\[
f(y_i | \mathbf{z}) = \exp \left( \frac{y_i^\prime \xi_i - b(\xi_i)}{a(\phi)} + c_i(y_i, \phi) \right),
\]

where \( \phi \) is a dispersion parameter and \( a(\cdot) \), \( b(\cdot) \), and \( c(\cdot) \) are known functions. \( \xi_i \) is associated with the conditional mean \( \mu_i = E(y_i | \mathbf{z}) \), which is associated with a linear predictor \( \eta_i = x_i^\prime \beta + z_i^\prime \alpha \) through a known link function \( g(\cdot) \) with \( g(\mu_i) = \eta_i \). According to the properties of the exponential family, one has \( b'(\xi_i) = \mu_i \). Under the canonical link, one has \( \xi_i = \eta_i \), that is, \( g = h^{-1} \) where \( h(\cdot) = b'(\cdot) \), \( h^{-1} \) represents the inverse function of \( h \).

2.2. Method of Simulated Moments (MSM)

MSM is introduced by McFadden (1989). MSM applies to cases where the theoretical moment function can’t be expressed as an analytic function of parameters because of the high-dimensional integrals involved. These moments have to be approximated by Monte Carlo simulation.
In the following, we describe the general results of MSM estimator for GLMM. Much of this notation is from Jiang (2009). Assume that the conditional density of $y_i$ given the vector of random effects $\mathbf{z}$ has the following form:

$$f(y_i | \mathbf{z}) = \exp((w_i/\phi)(y_i \xi_i - b(\xi_i)) + c_i(y_i, \phi)).$$

where $\phi$ is a dispersion parameter, and $w_i$'s are known weights with

$$w_i(x) = \begin{cases} 1, & \text{for ungrouped data} \\ n_i, & \text{for grouped data if the response is an average} \\ 1/n_i, & \text{response is a group sum.} \end{cases}$$

$b(\cdot)$ and $c_i(\cdot, \cdot)$ are known functions. For $\xi_i$, we assume a canonical link $\eta_i = \xi_i$.

Let $\mathbf{z} = (\mathbf{z}_1', \ldots, \mathbf{z}_q')$, where $\mathbf{z}_i$ is a $m_i \times 1$ random vector (with $m_1 + m_2 + \cdots + m_q = m$) whose components are independently distributed as $N(0, \sigma^2_i)$, $1 \leq r \leq q$. For convenience, let

$$\mathbf{z} = \mathbf{D}\mathbf{u},$$

where $\mathbf{D}$ is blockdiagonal with the diagonal blocks $\sigma_i\mathbf{I}_{m_i}$, $1 \leq r \leq q$, and $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_m)$.

Suppose the linear predictor associated with the link function is $X\mathbf{\beta} + Z\mathbf{z}$. $Z$ is an $n \times m$ design matrix. Let $Z_r$ be an $n \times m_r$ design matrix of random effects $\mathbf{z}_r$, so that $Z = (Z_1, \cdots, Z_q)$. For simplicity, we assume that $Z_r$, $1 \leq r \leq q$ are standard design matrices that each $Z_r$ consists only of 0s and 1s, and there is exactly one 1 in each row and at least one 1 in each column. We denote the $i$th row of $Z_r$ by $\mathbf{z}_{ir} = (z_{ir1}, \ldots, z_{irm})' \in \mathbb{R}^{m_r}$ with $1 \leq i \leq n$ and $1 \leq l \leq m_r$. We have $|z_{ir}|^2 = 1$ and for $s \neq t$, $\mathbf{z}_{is}\mathbf{z}_{tr} = 0$ or 1.

Let $\mathbf{X}_j$ be the $j$th column of design matrix $\mathbf{X}$. $\mathbf{W} = \text{diag}(w_i, 1 \leq i \leq n)$. Let $\psi = (\mathbf{u}'_1, \cdots, \mathbf{u}'_q)'$ with $\mathbf{u}_r = (u_{ir})_{1 \leq i \leq m_r}$, $\mathbf{u}_r \sim N(\mathbf{0}, \mathbf{I}_{m_r})$. $e(\psi, \mathbf{u}) = [b(\xi_{ir})]_{1 \leq i \leq n}$ with $\xi_{ir} = \sum_{j=1}^p x_{ij}\beta_j + \sum_{r=1}^q \sigma_r z_{ir} u_r$. Thus, the MM equations that do not involve $\phi$ are given by

$$\sum_{i=1}^n w_i x_{ij} y_i = \mathbf{X}_j \mathbf{W}(e(\psi, \mathbf{u})), 1 \leq j \leq p, \quad (3)$$

$$\sum_{(i,t) \in L} w_i w_j y_{it} = \mathbf{W}(e(\psi, \mathbf{u}) \mathbf{W}^2 e(\psi, \mathbf{u})), 1 \leq r \leq q, \quad (4)$$

where the expectations on the right-hand sides are with respect to $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_m)$. We approximate the right-hand sides by a simple Monte Carlo simulation. Let $\psi^{(1)}, \ldots, \psi^{(L)}$ be generated i.i.d. copies of $\psi$, the right side of (3) and (4) can be approximated by Monte Carlo averages $\mathbf{X}_j \mathbf{W} \big[ \frac{1}{L} \sum_{i=1}^L e(\psi^{(i)}, \mathbf{u}^{(i)}) \big], 1 \leq j \leq p$ and $\frac{1}{L} \sum_{i=1}^L e(\psi^{(i)}, \mathbf{u}^{(i)}) \mathbf{W}^2 e(\psi^{(i)}, \mathbf{u}^{(i)}), 1 \leq r \leq q$. 

The Standard Error of Method of Simulated Moment Estimator

3.
3. Variance Estimation of MSM Estimator Using Nonparametric Block Bootstrap

Parametric bootstrap is a simple way to estimate the standard errors. However, the estimated standard error is sensitive to the parameter estimators used to generate the simulated moments. In Sec. 3.1, we review nonparametric bootstrap and block bootstrap. In Sec. 3.2, we propose a two-stage block nonparametric bootstrap procedure for estimating standard error of MSM estimator for GLMM.

3.1. Bootstrap and Block Bootstrap

The basic idea of nonparametric bootstrap is to take resamples with replacement from the original sample. The resamples are then used to estimate the parameter of interest. The sampling variance is used to estimate the errors. Nonparametric bootstrap assumes that in drawing the resamples, it produces properties of the original sample.

When observations are correlated, resampling single observation that was considered by Efron (1979) for independent data failed to work. Block bootstrap was proposed by Kunsch (1989) for analyzing the time series data sets. By retaining the neighboring observations together within the blocks, the dependence structure of the random variables at short lag distances is preserved. Resampling blocks allows us to carry over this dependence information to the bootstrap variables. Liu and Singh (1992) independently suggested “Moving Block Bootstrap” (MBB). There are other proposed block bootstrap such as non overlapping block bootstrap (NBB) (Carlstein, 1986), circular block bootstrap (CBB) and the stationary bootstrap (SB) (Politis and Romano, 1992), the matched block bootstrap (MaBB) (Carlstein et al., 1998), and the tapped block bootstrap (TBB) (Paparoditis and Politis, 2001).

3.2. Nonparametric Block Bootstrap for GLMM

In this section, we first consider nonparametric block bootstrap for a simple logistic model. Then we extend the procedure to a general GLMM.

Consider a simple logistic model,

\[ \logit(P(Y_{ij} = 1)) = \mu + \alpha_i, \]

where \( 1 \leq i \leq m, 1 \leq j \leq k_i \) for each \( i \), and \( \alpha_i \)'s are i.i.d. normally distributed random variables with mean zero and variance \( \sigma^2 \). Here, \( m \) is the number of the factor levels and \( k_i \) is the number of observations within each factor level \( i \). For simplicity, we assume that \( k_i = k \), for \( i = 1, 2, \ldots, m \). Consider a sample:

\[
\begin{align*}
Y_1 & : Y_{11}, Y_{12}, \ldots, Y_{1k}, \\
Y_2 & : Y_{21}, Y_{22}, \ldots, Y_{2k}, \\
\vdots \\
Y_m & : Y_{m1}, Y_{m2}, \ldots, Y_{mk}.
\end{align*}
\]

The following gives a procedure of nonparametric block bootstrap standard error estimation for MSM estimators.
Step 1. Sample \( m \) numbers from 1 to \( m \) with replacement, say \( m_1, m_2, \ldots, m_m. \) Here, \( m_i \) may be equivalent to \( m_j \) even \( i \neq j. \) \( m_i \) is corresponding to the vector \( y_{m_i}. \)

Step 2. Sample \( k \) observations from \( y_{m_1}; k \) observations from \( y_{m_2}, \) etc. until \( k \) observations from \( y_{m_m} \) with replacement to form a block bootstrap sample.

Step 3. Calculate MSM estimators of \( \mu \) and \( \sigma^2 \) for the bootstrap sample. The MM estimating equations are as follows:

\[
\frac{1}{m} \sum_{i=1}^{m} y_i = E(y_i),
\]

and

\[
\frac{1}{m} \sum_{i=1}^{m} y_i^2 = E(y_i^2),
\]

where \( E(y_i) = kE(f(x)), \) \( E(y_i^2) = kE(f(x)) + k(k-1)E(f^2(x)). \) \( f(x) = \exp(\mu + \sigma x)/(1 + \exp(\mu + \sigma x)) \) and \( x \) is a standard normal random variable. Generate a sequence of standard normal random variables, \( x_1, x_2, \ldots, x_L. \) We approximate the expectations by

\[
E(f(x)) = \frac{\sum_{i=1}^{L} f(x_i)}{L}
\]

and

\[
E(f^2(x)) = \frac{\sum_{i=1}^{L} f^2(x_i)}{L}.
\]

The MSM estimators of \( \mu \) and \( \sigma \) can be solved from the system MM estimating equations.

Step 4. Repeat Steps 1 and 3 \( n \) times. Calculate the sampling variance of the MSM estimators from the \( n \) bootstrap samples.

Now we extend the above steps to a general GLMM case,

\[
g(u) = X\beta + Z\alpha.
\]

Rewrite (6) as

\[
g(u_i) = X_i^r\beta + Z_i^r\alpha,
\]

where \( X_i^r \) is the \( n_i \times p \) design matrix with \( X = (X_i^r)' \) and \( Z_i^r \) is the \( n_i \times m \) design matrix with \( Z = (Z_i^r)' \). Consider the sample

\[
y_1 : y_{11}, y_{12}, \ldots, y_{1n_1},
\]

\[
y_2 : y_{21}, y_{22}, \ldots, y_{2n_2},
\]

\[
\vdots
\]

\[
y_r : y_{r1}, y_{r2}, \ldots, y_{rn_r},
\]

\[
\vdots
\]

\[
y_m : y_{m1}, y_{m2}, \ldots, y_{mn_m},
\]

\[
\vdots
\]
where \( t \) is number of groups with \( \sum n_i = n, i = 1, 2, \ldots, t \). Continue with Step 1 to Step 4.

4. Simulation Studies

In this section we use simulation to study the performance of the nonparametric block bootstrap for estimating the standard error of MSM estimator. Simulation results are given in Table 1. Logistic normal model (5) is used in simulation study. Following are the simulation steps.

1. Set \( \mu = .2, \sigma = 1, m = 30 \) and \( k = 6 \). Generate a sample from (5).
2. Find MSM estimators \( \hat{\mu} \) and \( \hat{\sigma}^2 \). The replication number \( L \) for the simulated moments is set to 100.
3. Find nonparametric block bootstrap variance estimators \( BB[SE(\hat{\mu})] \) and \( BB[SE(\hat{\sigma}^2)] \) (refer to Sec. 3.2), the replication number is 200 for this step.
4. Repeat Steps (1)–(3) for \( n = 1,000 \) replications, record the average values of the MSM estimators, standard errors \( SE(\hat{\mu}), SE(\hat{\sigma}^2) \), and nonparametric block bootstrap standard errors of MSM estimators;
5. Repeat Steps (1)–(4) for different settings with \( m = 30, k = 20; m = 80, k = 6; \) and \( m = 80, k = 20 \). Table 1 gives the results.
6. Repeat Steps (1)–(5) for \( \mu = 1.0, \sigma = 1.0 \). Table 2 gives the results.

In Tables 1 and 2, \( \hat{\mu} \) and \( \hat{\sigma}^2 \) are the average values of the MSM estimators from the 1,000 replications. They are considered as the true mean and variance. \( SE(\hat{\mu}) \) and \( SE(\hat{\sigma}^2) \) are the average value of standard errors. They are considered as the true standard error estimates of \( \hat{\mu} \) and \( \hat{\sigma}^2 \). \( BB[SE(\hat{\mu})] \) and \( BB[SE(\hat{\sigma}^2)] \) are standard error estimates of \( \hat{\mu} \) and \( \hat{\sigma}^2 \) from nonparametric block bootstrap method. Numbers in parenthesis are the sample variances.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Simulation results: ( \mu = .2, \sigma = 1 ) in (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>( k )</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>80</td>
<td>6</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Simulation results: ( \mu = 1.0, \sigma = 1.0 ) in (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>( k )</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>80</td>
<td>6</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>
From Tables 1 and 2, it is clear that the method of simulated moments work well in estimating the $\mu$ and $\sigma^2$. Standard error estimates from nonparametric block bootstrap are close to the standard deviation.

5. Concluding Remarks

In this article, we studied standard error estimation of the method of simulated moment estimators for GLMM using nonparametric block bootstrap. We described a procedure to generate the nonparametric block bootstrap samples for use in a GLMM. The logistic model is used as an example in the simulation study to illustrate our proposed procedure. Simulation study indicates that nonparametric block bootstrap is doing well in estimating the standard error of MSM estimator for GLMM.

References