Approximate Estimation in Generalized Linear Mixed Models with Applications to the Rasch Model

M. L. FEDDAG
Department of Statistics
University of Warwick
Coventry CV4 7AL, U.K.
M.L.feddag@warwick.ac.uk

M. MESBAH
LSTA University of Paris 6
175 rue de Chevaleret, 75013 Paris, France
mesbah@ccr.jussieu.fr

Abstract—This article discusses two different approaches to estimate the difficulty parameters (fixed effects parameters) and the variance of latent traits (variance components) in the mixed Rasch model. The first one is the generalized estimating equations (GEE2) which uses an approximation of the marginal likelihood to derive the joint moments whilst the second approach uses the maximum of the approximate likelihood. We illustrate these methods with a simulation study and with an analysis of real data from a quality of life. © 2006 Elsevier Ltd. All rights reserved.

Keywords—Generalized linear mixed models, Cluster data, Fixed effects, Variance components, Generalized estimating equations (GEE), Approximate likelihood, IRT models, Rasch model, Quality of life.

1. INTRODUCTION

Generalized linear mixed models (GLMMs) are extensions of generalized linear models (GLMs) that accommodate correlated and overdispersed data by adding random effects to the linear predictor. Their broad applications are useful in various disciplines, such as the analysis of clustered data including longitudinal data or repeated measures. These models are useful for accommodating the overdispersion often observed among outcomes that nominally have binomial or Poisson distribution and for modelling the dependence among outcome variables in longitudinal or repeated measures designs. Such generalized linear mixed models are also increasingly used in various fields where subjective variables (called latent traits) need to be measured using questionnaires with polychotomous items. This is usual in health sciences and clinical trials, where these subjective variables could be pain, depression, or quality of life. Other examples come from marketing where satisfaction or attitudes need to be well measured and educational testing.

The authors gratefully acknowledge the guest editors and referees for their valuable comments and suggestions leading to the improvement of the paper. We also thank A. Leonard and V. Moquet for their help with the English representation.

0898-1221/06/$ - see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.camwa.2005.11.012

Typeset by A4AS-TEX
services where well-calibrated exams need to be produced. In such fields, the Rasch model, the
most popular IRT (item response theory) model, is very often used. Statistical properties of this
model are well known. We consider the Rasch model with random latent traits when the primary
interest is the population about the random sampling persons and/or the stochastic subject view.
This is also the case when the interest is the comparison of treatment groups. This mixed Rasch
model belongs to the family of logistic linear mixed models.

Estimation is complicated by the fact that these models are typically nonlinear with respect to
the random effects and have no closed form expressions for the marginal likelihood or indeed for
the marginal moments. Several approximate methods have been proposed. These include exact
integration-based methods using either numerical integration methods or Monte Carlo Markov
chain (MCMC) techniques. Such integration based techniques work extremely well when the
number of random effects is small. However, these techniques become increasingly more difficult
to use when the number of random effects increases and the computations are intensive. As an
alternative to numerical integration, two different first-order Taylor series expansion methods
have been used to approximate the marginal likelihood function and/or marginal moments. The
first expands the conditional mean or likelihood about the mean of the random effects and the
estimation is obtained with the generalized estimating equations (GEE), the approach defined
by Liang and Zeger [1] as an extension to the quasi-likelihood (see [2]). The second method
entails expanding the maximum of current estimates of the random effects, meaning the Laplace
approximation to the integrals. The estimation of different parameters is obtained by conditional
GEE for the fixed parameters and by the restricted estimation maximum likelihood (REML) or
the profiled maximum likelihood for the variance components.

This article considers two approximate methods of estimation in the mixed Rasch model and
investigates their properties. The first one is the estimating equations approach (GEE2) which
was previously proposed by Feddag et al. [3] to the logistic mixed models. This method involves
the approximations of the joint moments up to order four which are derived from the approximate
marginal likelihood meaning the Sutradhar and Rao [4] approximations. The estimators obtained
are consistent and asymptotically normal. The second approach is the maximum approximate
marginal likelihood (denoted MApL).

The paper is organized as follows. The generalized linear mixed model and a review on esti-
mation methods are given in Section 2. Thereafter in Section 3, we give approximations for the
marginal likelihood to the mixed Rasch model and we derive the approximate joint moments of
the variables. Next to this, we present two methods to estimate the difficulty parameters and
the variance component of the latent traits. In Section 4, we present some simulation results for
these methods and give an application using real data from a quality of life experiment. A brief
conclusion is presented in Section 5.

2. GENERALIZED LINEAR MIXED MODEL

Consider a sample of \( N \) independent random multivariate response \( y_i = (y_{i1}, \ldots, y_{in_i})' \), \( i = 1, \ldots, N \), where \( y_{ij} \) is the \( j \)th response to the \( i \)th cluster or subject. We shall assume that \( y_{ij} \)
depends on a \( p \times 1 \) vector of fixed covariates \( x_{ij} \) associated with a vector of fixed effect \( \beta = (\beta_1, \ldots, \beta_p)' \) and on a \( q \times 1 \) vector of fixed covariate \( z_{ij} \) associated with the multivariate \( q \times 1 \)
random effect \( b_i \). The generalized linear mixed model (GLMM) (see [5]) satisfies the following
conditions.

- Given \( b_i \), the variables \( y_{i1}, \ldots, y_{in_i} \) are mutually independent with a density function given by

\[
 f(y_{ij} | b_i, \beta) = \exp \left\{ \frac{y_{ij} \theta_{ij} - a(\theta_{ij})}{d_{ij}(\phi)} + c(y_{ij}, \phi) \right\},
\]

where \( \theta_{ij} \) is the canonical parameter and \( \phi \) is the scale parameter. The functions \( d_{ij} \) and \( c \)
are specific to each distribution.
The conditional mean and the conditional variance of $y_{ij}$ are given by

\[ E(y_{ij} \mid b_i) = \mu_{ij} = h^{-1}(x_{ij}' \beta + z_{ij}' b_i), \]

\[ \text{Var}(y_{ij} \mid b_i) = v(\mu_{ij} d_{ij}(q)), \]

where $h$ and $v$ are, respectively, the link and the variance function.

The random effects $b_1, \ldots, b_N$, are mutually independent with a common underlying distribution $G$ which depends on the unknown parameters $\alpha$.

We are interested in estimating $\beta, \alpha$. One approach is to leave $G$ completely unspecified and to use a nonparametric maximum likelihood estimation. One therefore often assumes $G$ to be of a specific parametric form, generally a multivariate normal with mean vector $0$ and covariance matrix $\Sigma = \Sigma(\alpha) = (\sigma_{ij})_{i,j=1,\ldots,q}$, and to use a parametric or semiparametric approach. In this case the marginal likelihood of $y = (y_1, \ldots, y_N)$ is given by

\[ L(\beta, \alpha \mid y) = (2\pi)^{-Nq/2} |\Sigma|^{-N/2} \prod_{i=1}^N \prod_{j=1}^{n_i} f(y_{ij} \mid b_i, \beta) \exp \left( -\frac{1}{2} b_i' \Sigma^{-1} b_i \right) \ db_i. \] 

The maximization of the above function is computationally difficult and requires evaluation of integrals where the integral’s dimension is equal to the number of random effects. Various methods have been proposed to circumvent this problem. These include exact integration-based methods using either numerical integration methods or Monte Carlo Markov chain (MCMC) techniques and approximations to the marginal likelihood or and marginal moments.

The numerical methods are centered on the adaptive quadrature [6]. An alternative to numerical integration is based on Monte Carlo Markov chain methods (see [7]). However, these techniques become increasingly more difficult to use when the number of random effects increases and the computations are intensive.

As an alternative to these methods, different approaches using approximations of the marginal likelihood or marginal moments have been proposed. The methods proposed by Schall [8] and Engel and Keen [9] are based on the transformation of GLMM to a classical linear mixed model (CLMM). The link function $h(.)$ applied to the data $y_i, i = 1, \ldots, N$ is linearized to the first order as given in [5],

\[ h(y_i) = h(\mu_i) + (y_i - \mu_i) h'(\mu_i) = y_i. \]

We obtain therefore a linear random effects model for the variables $y_i^{*}, i = 1, \ldots, N$. The estimation of the parameters is obtained by classical maximum likelihood or restricted maximum likelihood (REML). Lavergne and Trottier [10] have compared these methods with the approach proposed by Gilmour et al. [11]. Zeger et al. [12] have used an approximate mean vector and a working covariance matrix meaning an expansion to the link function $h(.)$ in a Taylor series about $b_i = 0$, to estimate the regression parameters by GEE approach. Breslow and Clayton [13] used the same approximations in the marginal quasi-likelihood (MQL) approach. The regression parameters are estimated as in [12] while the variance components are estimated by the REML or the profiled maximum likelihood. The second approximate method of Breslow and Clayton [13], namely, the penalized quasi-likelihood (PQL), is based on first-order Taylor expansions around the maximum of current estimates of the random effects via the first-order Laplace approximations to the integrals. These approaches produce biased estimates for both the regression and variance components parameters. Breslow and Lin [14] provided a correction factor for the estimates of the univariate variance components derived from the second-order Laplace approximations. Lin and Breslow [15] extend this bias correction to the GLMM with multivariate random effects. Many of these methods are implemented in the SAS macro GLIMMIX and in Splus macro GEEX. Feddag et al. [3] have used GEE2 to estimate simultaneously the fixed effects parameter and the variance components for the logistic mixed models. The estimators obtained are consistent.
and asymptotically normal. This method produces unbiased estimates for the regression effects and small bias for the variance components larger than 0.5. This approach is extended to the longitudinal mixed Rasch model (see [16]).

3. MIXED RASCH MODEL

3.1. Model

IRT (item response theory) models first appeared in the field of psychometry and educational sciences to quantify human behavior. They are now increasingly used in medicine to study psychological traits in psychiatry and, more recently, to assess quality of life in clinical trials or epidemiology. Generally, the quality of life of the patients is evaluated using questionnaires with dichotomous items. One of the most popular IRT models is the Rasch model (see [17]). To take into account the random variation in our application, we consider the mixed Rasch model with univariate random effects. These random effects can be regarded as additional error terms, to account for the correlation among observations within the same subject and for the heterogeneity between individuals. This model is a particular case of the GLMM defined in Section 2 where the link and variance function are, respectively, defined by

\[ h(t) = \ln \left( \frac{t}{(1-t)} \right), \quad v(t) = t(1-t), \]

and the covariates associated with the fixed effects and with the random effects are, respectively, given for all \( i, j \), by

\[ x_{ij} = (0, \ldots, 0, -1, 0, \ldots, 0), \quad z_{ij} = 1. \]

Consider a set of \( N \) individuals having answered a questionnaire of \( J \) dichotomous items. Let \( y_{ij} \) be the answers of individual \( i \) to item \( j \), where \( y_{ij} = 1 \) if the \( i^{th} \) individual has a positive response (correct, agree) for item \( j \) and \( y_{ij} = 0 \) (false, disagree) otherwise. The probability \( p_{ij} \) of the response of the \( i^{th} \) individual to the \( j^{th} \) item is given by

\[ p_{ij} = \frac{\exp [(b_i - \beta_j)y_{ij}]}{1 + \exp (b_i - \beta_j)}, \tag{3} \]

where \( \beta_j \) is the difficulty parameter to item \( j \) and \( b_i \) is the random variable called latent variable associated to subject \( i \).

The marginal likelihood of this model is given by

\[ L(\beta, \sigma^2 | y) = \frac{1}{(2\pi \sigma^2)^N} \prod_{i=1}^{N} \left\{ \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \frac{\exp [(b_i - \beta_j)y_{ij}]}{1 + \exp (b_i - \beta_j)} \exp \left( -\frac{b_i^2}{2\sigma^2} \right) db_i \right\}. \tag{4} \]

The maximization of the above function is computationally difficult and requires iterative techniques. Most researchers now use the EM algorithm. The integrals at two steps of the algorithm, the E step and the M step, are often approximated by the Gauss-Hermite quadrature (see [18]), a method with a slow rate of convergence. To avoid these difficulties, we propose the GEE2 approach, which turns out to be computationally less intensive and we compare it to the maximum approximate likelihood approach.

3.2. Approximations of Marginal Likelihood and Joint Moments

The aim of this section is to give the approximations of the joint moments up to order four of the observed variable \( y_i \), which we shall use later on in the GEE. Their computation requires an approximation of the marginal likelihood of \( y = (y_1, \ldots, y_N) \), given in equation (4). Under the following assumption:

\[ E(b_i^{2r}) = 0 \text{ for } r \geq 3, \tag{5} \]

expanding the conditional distribution of \( y_{ij}, i = 1, \ldots, N, j = 1, \ldots, J \), given by expression (3) in a Taylor series about \( b_i = 0 \), up to order four, we obtain the following theorem.
THEOREM 1. (See [3].)

1. Up to a term of order $O(\sigma^6)$, the likelihood $L(\beta, \sigma^2 \mid y)$ defined in (4) is approximated by

$$L^* (\beta, \sigma^2 \mid y) = \prod_{i=1}^{N} L^*_i (\beta, \sigma^2 \mid y),$$

with

$$L^*_i (\beta, \sigma^2 \mid y) = \left(1 + \frac{\sigma^2}{2} (A^2_i - B) + \frac{\sigma^4}{8} Q_i \right) \prod_{j=1}^{J} f_i^*(y_{ij} \mid \beta).$$

2. For all $i, j, m$ satisfying $1 \leq i \leq N, 1 \leq j \leq J, m = 2, 3, 4$, we have the approximations

(a) $E(y_{ij}) = \mu_j + O(\sigma^6)$, where

$$\mu_j = 1 + e^{\beta_j} + \frac{\sigma^2 e^{\beta_j} (e^{\beta_j} - 1)}{2(1 + e^{\beta_j})^3} + \frac{\sigma^4 e^{2\beta_j} (e^{2\beta_j} - 11e^{\beta_j} + 11e^{\beta_j} - 1)}{(1 + e^{\beta_j})^5},$$

(b) $E(\prod_{i=1}^{m} y_{ij}) = \zeta_{j_1 \ldots j_m} + O(\sigma^6)$, where

$$\zeta_{j_1 \ldots j_m} = \left( \prod_{i=1}^{m} \frac{1}{1 + e^{\beta_{ij}}} \right) \times \left[ 1 + \frac{\sigma^2}{2} (A^2_{j_1 \ldots j_m} - B_{j_1 \ldots j_m}) + \frac{\sigma^4}{8} Q_{j_1 \ldots j_m} \right].$$

The quantities involved in this theorem are defined below:

$$f_i^*(y_{ij} \mid \beta) = \exp \left[ -y_{ij} \beta_j - \ln (1 + e^{-\beta_j}) \right], \quad a_j = \ln (1 + \exp (-\beta_j)),
$$

$$A_{ij} = (y_{ij} - a_j^{(1)}), \quad B_j = a_j^{(2)}, \quad C_j = a_j^{(3)}, \quad F_j = a_j^{(4)}, \quad \hat{A}_j = \left(1 - a_j^{(1)}\right),
$$

$$A_i = \sum_{j=1}^{J} A_{ij}, \quad B = \sum_{j=1}^{J} B_j, \quad C = \sum_{j=1}^{J} C_j, \quad F = \sum_{j=1}^{J} F_j,
$$

$$Q_i = A_i^4 - 6A_i^2 B - 4A_i C + 3B^2 - F,
$$

$$\hat{A}_{j_1 \ldots j_m} = \sum_{l=1}^{m} \hat{A}_{j_l}, \quad B_{j_1 \ldots j_m} = \sum_{l=1}^{m} B_{j_l}, \quad C_{j_1 \ldots j_m} = \sum_{l=1}^{m} C_{j_l}, \quad F_{j_1 \ldots j_m} = \sum_{l=1}^{m} F_{j_l},
$$

$$\hat{Q}_{j_1 \ldots j_m} = \hat{A}_{j_1 \ldots j_m}^4 - 6\hat{A}_{j_1 \ldots j_m}^2 B_{j_1 \ldots j_m} - 4\hat{A}_{j_1 \ldots j_m} C_{j_1 \ldots j_m} + 3B_{j_1 \ldots j_m}^2 - F_{j_1 \ldots j_m},$$

and $a_j^{(t)}$ is the $t$-order derivative of $a_j$, $t = 1, \ldots, 4$.

3.3. Generalized Estimating Equations (GEE2)

Consider $y_i = (y_{i1}, \ldots, y_{iJ})'$, $i = 1, \ldots, N$, to be the outcomes of the mixed Rasch model defined above. Let $s_i$ be an $n_i(n_i - 1)/2 \times 1$ vector of empirical pairwise covariances defined by

$$s_i = (s_{i,j})'_{1 \leq j < i \leq J}, \quad \text{where} \quad s_{i,j} = (y_{ij} - \mu_{ij})(y_{il} - \mu_{il}).$$

Our approach (see [3]) for estimating parameters $\beta$ and $\sigma^2$ is as follows. Along with the basic estimating equations for the mean of $y_i, i = 1, \ldots, N$, which give estimators for the difficulty parameter $\beta$, we have used supplementary equations to estimate the parameter $\sigma$. These equations are based on the empirical covariances given by equation (9), which are unbiased estimators of the true covariance of the vector $y_i$.

The parameters $\beta$ and $\sigma^2$ are estimated by the solution of the generalized estimating equations given by

$$U (\beta, \sigma^2) = D'V^{-1} \sum_{i=1}^{N} \xi_i = 0,$$
where

\[ D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}, \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad \xi = \begin{pmatrix} y_i - \mu \\ s_i - \eta \end{pmatrix}, \]

with

\[ \mu = E(y_i) = (\mu_j)_{j=1,\ldots,J}, \quad \eta = E(s_i), \quad V_{11} = \text{Var}(y_i), \quad V_{12} = \text{Cov}(y_i, s_i), \]

\[ V_{22} = \text{Var}(s_i), \quad V_{21} = \text{Cov}(s_i, y_i), \quad D_{11} = \frac{\partial \mu}{\partial \beta} \text{ is } J \times J \text{ diagonal matrix}, \]

\[ D_{12} = \frac{\partial \mu}{\partial \sigma^2} \text{ is } J \times 1 \text{ vector}, \quad D_{22} = \frac{\partial \eta}{\partial \sigma^2} \text{ is } J(J-1)/2 \times 1 \text{ vector}, \]

\[ D_{21} = \frac{\partial \eta}{\partial \beta} \text{ is } J(J-1)/2 \times J \text{ matrix}. \]

All these quantities are computed using expressions (7) and (8) of Theorem 1.

The solution of (10), that is, \((\hat{\beta}, \hat{\sigma}^2)\), may be obtained by the Fisher scoring algorithm. The iterative procedure at step \((j + 1)\) is given by

\[
\begin{pmatrix} \hat{\beta}^{(j+1)} \\ \hat{\sigma}^{2(j+1)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^{(j)} \\ \hat{\sigma}^{2(j)} \end{pmatrix} + \frac{1}{N^2} \left( D' \hat{V}^{-1} \hat{D} \right)^{-1} \left( D' \hat{V}^{-1} \sum_{i=1}^{N} \hat{\xi}_i \right),
\]

where \(\hat{D}, \hat{V}, \hat{\xi}_i\) are, respectively, the values of \(D, V, \) and \(\xi_i\) at \((\hat{\beta}^{(j)}, \hat{\sigma}^{2(j)})\). And under some mild regularity conditions, \(N^{1/2}((\hat{\beta} - \beta)', (\hat{\sigma}^2 - \sigma^2)')\) is asymptotically multivariate normal with zero mean vector and covariance matrix which may be consistently estimated by

\[ \hat{V} = \lim_{N \to \infty} \frac{1}{N} \left( D' \hat{V}^{-1} \hat{D} \right)^{-1} \left( D' \hat{V}^{-1} \sum_{i=1}^{N} \hat{\xi}_i \hat{\xi}_i' \right) \left( D' \hat{V}^{-1} \hat{D} \right)^{-1}, \]

where \(\hat{\xi}_i, \hat{D}, \) and \(\hat{V}\) are, respectively, the values of \(\xi_i, D, \) and \(V\) at \((\hat{\beta}, \hat{\sigma}^2)\).

### 3.4. Maximum of the Approximate Marginal Likelihood

Now, we will estimate the parameters \((\beta, \sigma^2)\) by the maximum of the approximate likelihood (MApL) given by expression (6). The log-likelihood function is given by

\[
\ell^* = \ell^* (\beta, \sigma^2 \mid y) = \sum_{i=1}^{N} \sum_{j=1}^{J} \left( -y_{ij} \beta_j - \ln \left( 1 + e^{-\beta_j} \right) \right) + \sum_{i=1}^{N} \ln \left( 1 + \frac{\sigma^2}{2} (A_i^2 - B) + \frac{\sigma^4}{8} Q_i \right).
\]

Then the likelihood equations can be written

\[
\frac{\partial \ell^*}{\partial \beta} = \left( \frac{\partial \ell^*}{\partial \beta_j} \right)_{j=1,\ldots,J} = 0_J,
\]

\[
\frac{\partial \ell^*}{\partial \sigma^2} = 0,
\]

where \(0_J\) is the vector of zeros of order \(J\).

Thus, \((\hat{\beta}, \hat{\sigma}^2)\) is the solution to the score equations given by

\[
\frac{\partial \ell^*}{\partial \beta_j} = \sum_{i=1}^{N} \left( -y_{ij} + \frac{1}{1 + e^{\beta_j}} + \frac{H_i(\beta, \sigma)}{1 + (\sigma^2/2) (A_i^2 - B) + (\sigma^4/8) Q_i} \right) = 0, \quad j = 1, \ldots, J,
\]

\[
\frac{\partial \ell^*}{\partial \sigma^2} = \sum_{i=1}^{N} \left( (1/2) (A_i^2 - B) + (\sigma^2/4) Q_i \right)_{1 + (\sigma^2/2) (A_i^2 - B) + (\sigma^4/8) Q_i}, \quad (14)
\]
where

\[ H_i(\beta, \sigma) = \sigma^2 \left[ \frac{e^{\beta_j}}{(1 + e^{\beta_j})^2} + \frac{e^{\beta_j} (e^{\beta_j} - 1)}{2(1 + e^{\beta_j})^3} \right] + \frac{\sigma^4}{4} \left[ 2A_i^2 \frac{e^{\beta_j}}{(1 + e^{\beta_j})^2} - 6A_i B \frac{e^{\beta_j}}{(1 + e^{\beta_j})^2} - 3A_i e^{\beta_j} \frac{(1 - e^{\beta_j})}{(1 + e^{\beta_j})^3} - 2A_i e^{3\beta_j} + 4e^{\beta_j} - 1 - 2C \frac{e^{\beta_j}}{(1 + e^{\beta_j})^4} \right] \]

with \( A_i, B, \) and \( C \) the quantities are given in Theorem 1.

The estimates of the parameters are obtained by the Newton-Raphson algorithm, where the iterative procedure at step \((j + 1)\) is given by

\[
\begin{pmatrix}
\hat{\beta}^{(j+1)} \\
\hat{\sigma}^2(j+1)
\end{pmatrix} = \begin{pmatrix}
\hat{\beta}^{(j)} \\
\hat{\sigma}^2(j)
\end{pmatrix} - \begin{pmatrix}
\frac{\partial^2 \ell^*}{\partial \beta^2} & \frac{\partial^2 \ell^*}{\partial (\beta \sigma^2)} \\
\frac{\partial^2 \ell^*}{\partial (\sigma^2 \beta)} & \frac{\partial^2 \ell^*}{\partial (\sigma^2)^2}
\end{pmatrix}^{-1} \begin{pmatrix}
\frac{\partial \ell^*}{\partial \beta} \\
\frac{\partial \ell^*}{\partial \sigma^2}
\end{pmatrix} \begin{pmatrix}
(\hat{\beta}^{(j)}, \hat{\sigma}^2(j)) \\
(\hat{\beta}^{(j)}, \hat{\sigma}^2(j))
\end{pmatrix} .
\]

The covariance matrix of \((\hat{\beta}, \hat{\sigma}^2)\), \(W\) will be estimated by the inverse of the estimated information matrix

\[
\text{Inf} = -\mathbb{E} \begin{pmatrix}
\frac{\partial^2 \ell^*}{\partial \beta^2} & \frac{\partial^2 \ell^*}{\partial (\beta \sigma^2)} \\
\frac{\partial^2 \ell^*}{\partial (\sigma^2 \beta)} & \frac{\partial^2 \ell^*}{\partial (\sigma^2)^2}
\end{pmatrix} (\hat{\beta}, \hat{\sigma}^2) .
\]

4. ILLUSTRATIONS

4.1. Simulations

In this section we studied the sample performance to estimate the difficulty parameters \( \beta \) and the variance components \( \sigma^2 \) by the two approaches described in Section 3. We performed 200 replications with the same parameters considered in [19]: \( \beta = (-1, -0.5, 0.5, 1) \) and four values of \( \sigma^2: 0.2, 0.4, 0.6, 0.9 \). We considered three choices of the covariance matrix \( V \) in the GEE2 approach. The first one is \( V = V(4) \) where \( V(4) \) is the completely specified matrix (the joint moments up to order four are used). The second choice is \( V = V(3) \), where \( V(3) \) is the matrix corresponding to \( V(4) \), except that \( \text{Cov}(s_{i,j}, s_{i,k}) = 0 \), for all \((j, l) \neq (k, m)\). Finally the last choice is \( V = V(2) \), where \( V(2) \) is considered as diagonal and the elements \( \text{Cov}(y_{i,j}, s_{i,j}) \) of the matrix \( V(2) \) are zeros for all \( j \neq k, l \). The software package Splus is used for all the simulations. The GEE2 approach estimation procedure successfully converged to estimate in each replication for small \( \sigma^2 \). In contrast, this method encountered some estimation problems for large values of \( \sigma^2 (\sigma^2 = 0.9) \). For all values of \( \sigma^2 \), the method of maximum likelihood encountered some problems in converging. The percentage of nonconverged replications is higher than the GEE2 approach under the matrix \( V(4) \). We present in Table 1 the mean and the standard error (denoted s.e. and given in brackets) of the 200 estimates values obtained for the difficulty parameters \( \beta \) and variance components \( \sigma^2 \).

From Table 1, we note that the largest bias for the difficulty parameter estimates \( \beta_j \), \( j = 1, \ldots, 4 \), is equal to two percent for the different approaches. Their standard errors \( \text{s.e.}(\hat{\beta}_j) \) have the same order of error and the mutual differences are less than one percent. Overall, the results of the simulation show that the GEE2 approach under all structures of covariance matrices have some bias for the variance components estimates. This bias is less than 1.5 percent for \( \sigma^2 = 0.2 \) and less than 4.3 percent for \( \sigma^2 = 0.4 \). We note that this bias is large under \( V(4) \) relative to the two other
Table 1. Simulations results for the Rasch model with difficulty parameters \( \beta = (-1, -0.5, 0.5, 1) \) and variance components \( \sigma^2 = 0.2, 0.4, 0.6, 0.9 \) for \( N = 500, J = 4 \).

<table>
<thead>
<tr>
<th>( \sigma^2 )</th>
<th>Approach</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>GEE(V(4))</td>
<td>-1.001 (0.104)</td>
<td>-0.498 (0.094)</td>
<td>0.498 (0.099)</td>
<td>1.000 (0.106)</td>
<td>0.215 (0.130)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(3))</td>
<td>-1.003 (0.106)</td>
<td>-0.498 (0.100)</td>
<td>0.504 (0.096)</td>
<td>1.000 (0.103)</td>
<td>0.214 (0.106)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(2))</td>
<td>-1.003 (0.106)</td>
<td>-0.498 (0.100)</td>
<td>0.504 (0.096)</td>
<td>1.000 (0.103)</td>
<td>0.213 (0.105)</td>
</tr>
<tr>
<td></td>
<td>MApL</td>
<td>-0.993 (0.104)</td>
<td>-0.493 (0.102)</td>
<td>0.507 (0.092)</td>
<td>1.002 (0.115)</td>
<td>0.204 (0.101)</td>
</tr>
<tr>
<td>0.4</td>
<td>GEE(V(4))</td>
<td>-0.990 (0.107)</td>
<td>-0.489 (0.097)</td>
<td>0.509 (0.098)</td>
<td>1.011 (0.103)</td>
<td>0.443 (0.238)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(3))</td>
<td>-1.006 (0.109)</td>
<td>-0.501 (0.101)</td>
<td>0.494 (0.106)</td>
<td>1.000 (0.108)</td>
<td>0.413 (0.140)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(2))</td>
<td>-0.999 (0.110)</td>
<td>-0.494 (0.104)</td>
<td>0.502 (0.102)</td>
<td>1.000 (0.109)</td>
<td>0.415 (0.143)</td>
</tr>
<tr>
<td></td>
<td>MApL</td>
<td>-1.006 (0.092)</td>
<td>-0.507 (0.106)</td>
<td>0.498 (0.089)</td>
<td>1.010 (0.109)</td>
<td>0.398 (0.126)</td>
</tr>
<tr>
<td>0.6</td>
<td>GEE(V(4))</td>
<td>-0.983 (0.114)</td>
<td>-0.474 (0.101)</td>
<td>0.508 (0.101)</td>
<td>1.006 (0.109)</td>
<td>0.682 (0.300)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(3))</td>
<td>-1.001 (0.115)</td>
<td>-0.492 (0.099)</td>
<td>0.495 (0.103)</td>
<td>0.997 (0.113)</td>
<td>0.654 (0.212)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(2))</td>
<td>-1.001 (0.115)</td>
<td>-0.493 (0.098)</td>
<td>0.495 (0.103)</td>
<td>0.997 (0.113)</td>
<td>0.647 (0.189)</td>
</tr>
<tr>
<td></td>
<td>MApL</td>
<td>-0.983 (0.113)</td>
<td>-0.483 (0.111)</td>
<td>0.505 (0.099)</td>
<td>0.994 (0.124)</td>
<td>0.570 (0.131)</td>
</tr>
<tr>
<td>0.9</td>
<td>GEE(V(4))</td>
<td>-0.977 (0.114)</td>
<td>-0.476 (0.101)</td>
<td>0.499 (0.101)</td>
<td>1.015 (0.109)</td>
<td>0.987 (0.300)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(3))</td>
<td>-1.001 (0.118)</td>
<td>-0.488 (0.106)</td>
<td>0.485 (0.108)</td>
<td>0.992 (0.119)</td>
<td>0.996 (0.272)</td>
</tr>
<tr>
<td></td>
<td>GEE(V(2))</td>
<td>-1.000 (0.118)</td>
<td>-0.489 (0.104)</td>
<td>0.487 (0.109)</td>
<td>0.990 (0.118)</td>
<td>0.997 (0.250)</td>
</tr>
<tr>
<td></td>
<td>MApL</td>
<td>-0.991 (0.113)</td>
<td>-0.487 (0.106)</td>
<td>0.482 (0.109)</td>
<td>0.975 (0.118)</td>
<td>0.744 (0.102)</td>
</tr>
</tbody>
</table>

Covariance structures. This bias is negligible for small values of \( \sigma^2 \) in the maximum approximate likelihood approach. This bias tends to increase as the values of the variance components increase. The results for the standard error (s.e.(\( \hat{\beta}^2 \))) are similar. We note that the GEE2 approach is computationally less intensive than the MApL method.

4.2. Example

In this section, we illustrate the application of these two approaches to the analysis of real data from a quality of life. The sample is composed of 470 depressive patients who answered the French version of the emotional behavior subscale of sickness impact profile (SIP) questionnaire (see [20] for the international version). This questionnaire includes 12 dimensions (subscales), each one relating to a particular aspect of the quality of life. The content of items and their frequencies are presented in [3].

Table 2 presents the estimation of difficulty parameters \( \hat{\beta} \) and variance components \( \hat{\sigma}^2 \), and their standard errors (s.e.). The standard error is estimated using expression (12) for the GEE2 approach and expression (16) for the MApL approach. We also present the results of the marginal maximum likelihood (MML) and the conditional maximum likelihood (CML) with the use of the software program RSP (see [21]). The MML approach uses the EM algorithm where the integrals in the two steps of the algorithm are approximated by the Gauss-Hermite quadrature.

From Table 2, we note that the estimates obtained under the different approaches are similar. The large difference between the difficulty parameter items is very small. It is clear that the most difficult item is number 4 (with the largest estimation) and the easiest is item 7 (with a smaller estimation). The estimates of the variance component are very close and the smaller one is given by the MApL approach. The standard errors of all estimates are very close.

5. CONCLUSION

We have proposed a statistical estimation approach for a mixed Rasch model, based on the maximum of the approximate marginal likelihood. This approach is compared to the GEE2 method, previously proposed by Feddag et al. [3] to the logistic mixed models. These two approaches are based on the approximate likelihood to estimate simultaneously the difficulty parameters and the variance of the latent trait for the mixed Rasch model. The simulation results show that
Approximate Estimation in Generalized Linear Mixed Models

Table 2. Parameter estimates for \((\beta, \sigma^2)\) and their standard errors (s.e.) of the SIP data \((N = 470, J = 9)\).

<table>
<thead>
<tr>
<th>Approach</th>
<th>GEE((V(4)))</th>
<th>GEE((V(3)))</th>
<th>GEE((V(2)))</th>
<th>MApL</th>
<th>MML</th>
<th>CML</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>-0.707</td>
<td>-0.702</td>
<td>-0.702</td>
<td>-0.700</td>
<td>-0.683</td>
<td>-0.708</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>-0.270</td>
<td>-0.246</td>
<td>-0.240</td>
<td>-0.240</td>
<td>-0.273</td>
<td>-0.268</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>0.096</td>
<td>0.128</td>
<td>0.104</td>
<td>0.123</td>
<td>0.099</td>
<td>0.105</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>2.050</td>
<td>2.028</td>
<td>2.028</td>
<td>1.232</td>
<td>2.037</td>
<td>1.997</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>-1.269</td>
<td>-1.221</td>
<td>-1.216</td>
<td>-1.216</td>
<td>-1.270</td>
<td>-1.266</td>
</tr>
<tr>
<td>(\beta_6)</td>
<td>0.838</td>
<td>0.863</td>
<td>0.855</td>
<td>0.863</td>
<td>0.846</td>
<td>0.844</td>
</tr>
<tr>
<td>(\beta_7)</td>
<td>-1.418</td>
<td>-1.411</td>
<td>-1.411</td>
<td>-1.411</td>
<td>-1.415</td>
<td>-1.411</td>
</tr>
<tr>
<td>(\beta_8)</td>
<td>-1.376</td>
<td>-1.371</td>
<td>-1.371</td>
<td>-1.371</td>
<td>-1.375</td>
<td>-1.371</td>
</tr>
<tr>
<td>(\beta_9)</td>
<td>-0.601</td>
<td>-0.570</td>
<td>-0.570</td>
<td>-0.570</td>
<td>-0.608</td>
<td>-0.603</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.703</td>
<td>0.674</td>
<td>0.482</td>
<td>0.482</td>
<td>0.678</td>
<td>0.603</td>
</tr>
</tbody>
</table>

estimators of the difficulty parameters are unbiased. The bias for the estimates of the variance component is negligible for small values of \(\sigma^2\). We note also that the GEE2 method is less intensive than the maximum approximate likelihood approach. We present an example of real data from a quality of life in which we compare these two approaches with the methods of marginal and conditional maximum likelihood.

REFERENCES

