International Journal of Probability and Statistics 2012, 1(3): 67-73
DOI: 10.5923/j.ijps.20120103.03

Generalized Linear Mixed Models for Longitudinal Data

Ahmed M. Gad*, Rasha B. El Kholy

Department of Statistics, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt

Abstract The study of longitudinal data plays a significant role in medicine, epidemiology and social sciences. Typically, the interest is in the dependence of an outcome variable on the covariates. The Generalized Linear Models (GLMs) were proposed to unify the regression approach for a wide variety of discrete and continuous longitudinal data. The responses (outcomes) in longitudinal data are usually correlated. Hence, we need to use an extension of the GLMs that account for such correlation. This can be done by inclusion of random effects in the linear predictor; that is the Generalized Linear Mixed Models (GLMMs) (also called random effects models). The maximum likelihood estimates (MLE) are obtained for the regression parameters of a logit model, when the traditional assumption of normal random effects is relaxed. In this case a more convenient distribution, such as the lognormal distribution, is used. However, adding non-normal random effects to the GLMM considerably complicates the likelihood estimation. So, the direct numerical evaluation techniques (such as Newton - Raphson) become analytically and computationally tedious. To overcome such problems, we propose and develop a Monte Carlo EM (MCEM) algorithm, to obtain the maximum likelihood estimates. The proposed method is illustrated using a simulated data.

Keywords Generalized Linear Mixed Models, Logistic Regression, Longitudinal Data, Monte Carlo EM Algorithm, Random Effects Model

1. Introduction

Longitudinal data consist of repeated observations, for the same subject, of an outcome variable. There may be a set of covariates for each subjects. Let $\boldsymbol{y_i} = (y_{i1}, y_{i2}, \dots, y_{imi})^{\mathrm{T}}$ be an m_i x 1 vector representing the observed sequence of the outcome variable Y_{it} recorded at time t = 1,2, ... , m_i, for the *i*th subject, *i=1,2, ..., n*. Also, assume that $\mathbf{x}_{ij} = (x_{it1}, x_{it2}, \dots, x_{itp})$ is an *1x p* vector of *p* covariates observed at time *t*. Thus, X_i is an $m_i x\ p$ matrix of covariates corresponding to the *i*th subject taking the form:

$$\begin{pmatrix} x_{i11} & \cdots & x_{i1p} \\ \vdots & \ddots & \vdots \\ x_{imi1} & \cdots & x_{imip} \end{pmatrix}.$$

For simplicity we can assume that m_i=*m*.

The primary focus is on the dependence of the outcome on the covariates[3,5]. In other words, we are usually interested in the inference about the regression coefficients, in the usual linear models of the form

$$Y_i = X_i\beta + \epsilon_{i,}$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is a p x1 vector of the regression coefficients. It is usually assumed that the errors, ϵ_i, are independent and identically normally distributed. Therefore, these models are not used in situations when response variables have distributions other than the normal, or even when they are qualitative rather than quantitative. Examples include binary longitudinal data.

To solve this problem, Reference[18] introduce the generalized linear models (GLM) as a unified framework to model all types of longitudinal data[13, 15, 24]. These models assume that the distribution of Y_{it} *(i = 1, ..., n, t = 1, 2, ..., m)* belongs to the exponential family. The exponential family distributions can be written in the form:

$$f(y_{it}, \theta_{it}) = e^{[(y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})]\emptyset}, \quad (1)$$

where $a(.)$ and $b(.)$ are specific functions defined according to the underlying distribution of the exponential family. Some examples of univariate distributions are given in Table (1). The parameters $\theta_{it} = h(\eta_{it})$ and $\eta_{it} = x_{it}\beta$. Also ϕ is the scale parameter and it is treated as a nuisance parameter when the main interest is in the regression coefficients[13]. The distribution in Equation (1) is the canonical form and θ_{it} is called the natural parameter of the distribution.

The first two moments of y_{it} in Equation (1) are given by

$$\mathrm{E}(y_{it}) = \mu_{it} = a'(\theta_{it}) \text{ and } \mathrm{Var}(y_{it}) = \frac{a''(\theta_{it})}{\emptyset} \quad (2)$$

$$\mathrm{b1} = -\frac{1}{2}(y^2 + \sigma^2 \ln(2\pi\sigma^2))$$

For the GLM, the linear combination of the parameters of interest, $\beta_1, \beta_2, \dots, \beta_p$, are equal to some function of the expected value of y_{it} (μ_{it}), i. e., $g(\mu_{it}) = \eta_{it}$ where g is a monotone and differentiable function called the link function. From Equation (1) and Equation (2), we can write the inverse of the link function as $g^{-1} = a'\,o\,h$. Note that a link function in which $\theta_{it} = \eta_{it}$ (e.g., h is the identity function)

* Corresponding author:
ahmed.gad@feps.edu.eg (Ahmed M. Gad)
Published online at http://journal.sapub.org/ijps

is called the canonical link function. It is sometimes preferred because it often leads to simple interpretable reparametrized models. By looking at θ in Table 1, we can see that the canonical link functions that correspond to the given distributions are the identity (normal distribution), the log (Poisson distribution) and the logit (binomial distribution) functions. Hence, the GLM has three elements:

Table 1. Some common univariate distributions of exponential family

Distribution	θ	*A*	*b*	ϕ
Normal (μ, σ^2)	μ	$\theta^2/2$	b1	$1/\sigma^2$
Poisson (λ)	ln λ	e^{θ}	-ln*(y!)*	1
Binomial (n, p)	$\ln(\frac{p}{1-p})$	$n \ln(1+e^{\theta})$	$\ln(n_{cy})$	1

1. The random component which is the response variables Y_{it} 's that are assumed to share the same distribution from the exponential family,

2. The systematic component which is the explanatory variables that produce the linear predictor $\eta_{it} = x_{it}\beta$, and

3. The link function which is a monotone and differentiable function g relating the mean μ_{it} and the linear predictor η_{it}.

The aim of this paper is to obtain the maximum likelihood estimates (MLE) of the regression parameters for the logit model relaxing the normality assumption. In this case the estimation process is cumbersome and intractable. Hence, MCMC techniques could be alternative choice. We propose and develop the Monte Carlo EM (MCEM) algorithm, to obtain the maximum likelihood estimates. The proposed method is illustrated using simulated data.

The rest of the paper is organized as follows. In Section 2, we review three distinct approaches to model longitudinal data. Also, we introduce the random effect model. Section 3 discusses several alternatives to maximum likelihood estimation for GLM. The EM algorithm and its variant, namely Monte Carlo EM (MCEM), are described and applied to the random effect model in Section 4. In Section 5, we implement the proposed MCEM algorithm to a simulated binary data. The results obtained from the simulated data are presented in Section 6. Finally, conclusions are given in Section 7.

2. Modeling Longitudinal Data

There are three distinct strategies for modeling longitudinal data. Each strategy provides different way to model the individual y_{it} in terms of x_{it} , taking into consideration the possible correlation between the subject's measurements.

2.1. The Marginal Model

In this approach two models are specified; one for the marginal mean, $\mathrm{E}(Y_i)$, and the other for the marginal covariance, $\mathrm{Cov}(Y_i)$. In other words, we assume a known structure for the correlation between a subject's measureme nts. The two more natural generalization of the diagonal covariance matrix (case of uncorrelated measurements) are the uniform and the exponential correlation structures. Modelling the mean and the covariance separately has the advantage of making valid inferences about β even when an incorrect form of the within-subject correlation structure is assumed.

2.2. The Transition Model

This approach combines the assumptions about the dependence of Y_i on $\boldsymbol{x_{it}}$ and the correlation among repeated Y_i's into a single equation. The idea behind this approach is that correlation within subject arises because one response is explicitly caused by others. We can specify a regression model for the conditional expectation, $\mathrm{E}(Y_{it}|H_{it})$, as an explicit function of $\boldsymbol{x_{it}}$ and the other observations on the same subject. The function H_{it} is a function of Y_{it}^{-}, where Y_{it}^{-} is the set of all observations on the subject *i* except at time *t*.

For longitudinal data, it is natural to reduce the set Y_{it}^{-} to include only observations prior to time *t*; $(Y_{i1}, \ldots, Y_{it-1})$. A well known example of the transition model is the first-order autoregressive model in which Y_{it} depends on the past history in Y_i only through the preceding measurement Y_{it-1}. Note that the covariance matrix for Y_{it} in this case corresponds to the marginal model with exponential correlation structure.

2.3. The Random Effects Model

This model assumes that the correlation among a subject's measurements arises from sharing unobserved variables. That is, there are random effects which represent unobserved factors that are common to all responses for a given subject. These random effects vary across subjects[8].

Using the GLM framework, we assume that, conditional on the unobserved variables (γ_i), we have independent responses from a distribution belongs the exponential family, i.e., $f(Y_{it}|\gamma_i)$ has the form in Equation (1) and the conditional moments are given by

$$E(y_{it}|\gamma_i) = a'(\theta_{it}) \text{ and } Var(y_{it}|\gamma_i) = \frac{a''(\theta_{it})}{\emptyset}.$$

The general specifications of the generalized linear mixed model (GLMM) are:

1. The conditional mean, $E(y_{it}|\gamma_i)$, is modelled by

$$g(E(Y_{it}|\gamma_i) = \boldsymbol{X}_{it}\beta + z_{it}\,\gamma_i,$$

where g is an appropriate known link function. The Y_{it}, $\boldsymbol{X}_{it}$ and β are the outcome variable, the covariates and the regression coefficients vector respectively. The $\boldsymbol{z}_{it} = (z_{i1}, \ldots, z_{iq})$ is an $1x\ q$ subset of $\boldsymbol{X}_{it}$ associated with random coefficients and γ_i is a $q\ x\ 1$ vector of random effects with mean *0* and variance *D*.

2. The random effects, γ_i, are independent realizations from a common distribution.

3. Given the actual coefficients, γ_i, for a given subject, the repeated observations for that subject are mutually independent.

Hence, the GLMM is an extension of GLM that includes random effects in the linear predictor, giving an explicit probability model that explains the origin of the correlations [9,10].

Note that for linear model, where the response variable has a normal distribution and the identity link function is used, the random effect model coincides with the marginal model assuming the uniform correlation structure.

3. Estimation of GLM

The method of maximum likelihood is the theoretical basis for parameter estimation in GLM. However, in longitudinal data framework, the presence of within subject correlation renders the use of standard maximum likelihood estimation methods to be problematic. The reason behind this problem is that there are no multivariate generalization to non-normal distributions. As a result, Reference[11] introduce the generalized estimating equations (GEE) method. This method does not make use of the underlying multivariate distribution. It uses certain information associated with the marginal distribution rather than the actual likelihood, which is referred to as quasi-likelihood method[1, 12, 13].

References[11] and[25] introduce a class of estimating equations to account for correlated measurements of longitudinal data. The GEEs can be thought of as an extension of quasi-likelihood to the case where the variance cannot be fully specified in terms of the mean, but rather additional correlation parameters must be estimated. Several alternative methods for analyzing longitudinal data can be implemented using the GEEs, assuming different designs or structures of the correlation matrix[26, 27]. It is difficult to adapt the GEEs for the GLMM. This is due to the fact that the GEEs work most naturally for models specified marginally, not for the GLMMs which are specified conditionally on the random effects[14]. The GEEs, by themselves, do not help to separate out different sources of variation. In addition, they are not direct technology for best prediction of random effects.

The penalized quasi-likelihood (PQL) is another alternative to the maximum likelihood estimation when correlated data are to be incorporated in the GLM. In this approach a penalty function is added to the quasi-likelihood to prevent the random effects from getting too “big”. This method works well for the GLMM when the conditional distribution of the data given the random effects is approximately normal. However, the method can fail badly for distributions that are far from normal[14]. One possible reason for this drawback is the large number of approximati ons needed to solve the integration of the log quasi-likeliho od with the additional “penalty” term.

As can be seen, the alternatives to ML fail to work well for many of the GLMMs. Therefore, the classical method remains favourable even though incorporating random factors in the linear predictor of the GLM leads to difficult-to-handle likelihoods. To overcome this difficulty, approaches for approximating or calculating and then maximizing the likelihood are explored.

4. Likelihood Function for GLMM

The GLM often leads to means which are non-linear in parameters, or models with non-normal errors. Also, it leads to missing data or dependence among the responses. This results in a non-quadratic likelihood function in the parameters. Hence, it gives rise to nonlinearity problems in ML estimation[16, 19, 23].

Recall the notation for the GLMM from Section 2.3. Let $y_i = (y_{i1}, y_{i2}, \dots, y_{im})^T$ denote the observed data vector of size *m* for the *i*th subject. The conditional distribution of y_{it} given γ_i (the random effect vector for the *i*th subject) follows a GLM of the form in Equation (1) with linear predictor $\eta_{it} = X_{it}\beta + z_{it}\ \gamma_i$. The vector $x_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})$ represents the *t*th row of X_i, the model matrix for the fixed effects, β. The $z_{it} = (z_{it1}, z_{it2}, \dots, z_{itq})$ represents the *t*th row of Z_i, the model matrix for random effects, γ_i, corresponding to the *i*th subject.

We now formulate the notion of a GLMM:

$$\begin{gathered} Y_{it}|\gamma_i \sim \text{indep } f(y_{it}|\gamma_i, \beta, \phi) \\ f(y_{it}|\gamma_i, \beta, \phi) = \exp[y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})]\phi \\ \mathrm{E}(Y_{it}|\gamma_i) = \mu_{it} \\ g(\mu_{it}) = \eta_{it} = \boldsymbol{x_{it}}\beta + \boldsymbol{z_{it}}\gamma_i \\ \gamma_i \sim K(\gamma_i|D), \end{gathered} \tag{4}$$

where *K(.)* is the density function of the random effect γ_i with variance-covariance matrix *D*.

The probability of any response pattern Y_i (of size *m*), conditional on the random effects γ_i is equal to the product of the probabilities of the observations on each subject, because they are independent given the common random effects.

$$f(y_i|\gamma_i, \beta) = \prod_{t=1}^{m} f(y_{it}|\gamma_i, \beta).$$

Thus the likelihood function for the parameter β and *D* can be written as:

$$L(\beta, D, \phi|y) = \prod_{i=1}^{n} \int f(y_i|\gamma_i, \beta, \phi)\, K(\gamma_i|D) d\,\gamma_i \tag{5}$$

The above likelihood is analytically intractable except for normal linear models. Traditionally, ML estimation, in these situations, is carried out using numerical iterative methods such as the Newton - Raphson (NR) method and the Fisher scoring method. Under reasonable assumptions of the likelihood and a sufficiently accurate starting value, the sequence of iterations produced by the NR method enjoys local quadratic convergence. This is regarded as a major strength of the NR method, for more details see[17].

However in applications, even fairly simple cases, these methods could be tedious analytically and computationally. The EM algorithm offers an attractive alternative for iterative ML estimation in a variety of settings involving missing data and incomplete information.

4.1. The EM Algorithm

The EM algorithm[2] is a method to obtain the ML estimates, in presence of incomplete data, which avoids an

explicit calculation of the observed data log-likelihood. The EM algorithm iterates two steps: the E-step and the M-step. In the E-step, the expected value of the log-likelihood of the complete data, given the observed data and the current parameter estimates, is obtained[22]. Thus the computation in the M-step can be easily applied to pseudo-complete data. The observed (incomplete) data likelihood function always increases or stays constant at each iteration of the EM algorithm. For more details see[17].

A typical assumption in the GLMM is to consider the random effect, γ_i, as missing data. The complete data is then (Y, γ_i) and the complete data likelihood is given by

$$L(\beta, D, \phi|y) = \prod_{i=1}^{n} \int f(y_i|\gamma_i, \beta, \phi)\, K(\gamma_i|D) d\,\gamma_i$$

Although the observed data likelihood function in Equation (5) is complicated, the complete data likelihood is relatively simple. In other words, the integration over the random effects in Equation (5) is avoided since the value of (the missing data) will be simulated during the EM algorithm and will be no longer unknown.

The log-likelihood is given by

$$\ln L(\beta, D, \phi|y, \gamma) = \sum_{i=1}^{n} \ln f(y_i|\gamma_i, \beta, \phi) + \ln K(\gamma_i|D) \quad (6)$$

The choice of γ to be the missing data has two advantages. First, upon knowing γ_i, the Y_{it}'s are independent. Second, in the M-step, where the maximization is with respect to the parameters β, ϕ and D, the parameters β and ϕ only in the first term of Equation (6). Thus, the M-step with respect to β and ϕ uses only the GLM portion of the likelihood function. So, it is similar to a standard GLM computation assuming that γ is known. Maximizing with respect to D, in the second term, is just maximum likelihood using the distribution of γ after replacing sufficient statistics (in the case that *K(.)* belongs to the exponential family) with the conditional expected values.

For the GLMM of the form in Equation (4), at the (r + 1) iteration, starting with initial values $\beta^{(0)}, \phi^{(0)}$ and $D^{(0)}$, the EM algorithm follows the steps:

1. The E-step (Expectation step)

Calculate the expectations with respect to the conditional distribution using the current parameters' value $\beta^{(r)}, \phi^{(r)}$ and $D^{(r)}$,

(a) $E\left[\ln f(y|\gamma, \beta, \phi)|y, \beta^{(r)}, \phi^{(r)}\right]$.

(b) $E\left[\ln K(\gamma|D)|y, \beta^{(r)}, \phi^{(r)}\right]$.

2. The M-step (Maximization step)

Find the values

(a) $\beta^{(r+1)}$ and $\phi^{(r+1)}$ that maximizes 1 (a).

(b) $D^{(r+1)}$ that maximizes 1 (b).

If convergence is achieved, then the current values are the MLEs, otherwise increment $r = r + 1$ and repeat the two steps.

4.2. The Monte Carlo EM Algorithm (MCEM)

In general, the expectations in the E-step above, cannot be obtained in a closed form. So we propose using the Monte Carlo Markov Chains (MCMC) techniques[6, 20, 21]. A random draw from the conditional distribution of $\gamma|y$ is obtained. Then the required expectations are evaluated via Monte Carlo approximations.

The Metropolis-Hastings algorithm can be implemented to draw a sample $\{\gamma^{(l)}; l = 1,2, \dots, L\}$ from the conditional distribution of $\gamma|y$. A candidate value γ^* is generated from a proposal distribution *Q(.)*. This potential value is accepted, as opposed to keeping the previous value, by a probability

$$\mathcal{A}(\gamma, \gamma^*) = \min[1, \frac{f(\gamma^*|y, \beta, \phi)Q(\gamma|\gamma^*)}{f(\gamma|y, \beta, \phi)Q(\gamma^*|\gamma)}]. \quad (7)$$

The proposal distribution can be chosen as *Q(.) = K(.)*, the density function of the random effect. This leads to a simplified form of the ratio in the probability of acceptance as

$$\frac{f(\gamma^*|y, \beta, \phi)Q(\gamma|\gamma^*)}{f(\gamma|y, \beta, \phi)Q(\gamma^*|\gamma)} = \frac{f(y|\gamma^*, \beta, \phi)K(\gamma^*)Q(\gamma|\gamma^*)}{f(y|\gamma, \beta, \phi)K(\gamma)Q(\gamma^*|\gamma)} = \frac{\prod_{i=1}^{n} f(y_{i|}\gamma_i^*, \beta, \phi)}{\prod_{i=1}^{n} f(y_{i|}\gamma_i, \beta, \phi)}$$

This calculation only involves the conditional distribution of $y|\gamma$.

Incorporating the Metropolis algorithm into the EM algorithm, starting from initial values $\beta^{(0)}, \phi^{(0)}$ and $D^{(0)}$, at iteration (r + 1),

1. Generate L values $\{\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(L)}\}$ from the conditional distribution $\gamma|y$ using the Metropolis algorithm and the current parameters' values $\beta^{(r)}, \phi^{(r)}$ and $D^{(r)}$.

2. The E-step (Expectation step)

Calculate the expectations as Monte Carlo estimates

(a) $E\left[\ln f(y|\gamma, \beta, \phi)|y, \beta^{(r)}, \phi^{(r)}\right] = \frac{1}{L}\sum_{l=1}^{L} \ln f(y|\gamma^{(l)}, \beta, \phi)$.

(b) $E\left[\ln K(\gamma|D)|y, \beta^{(r)}, \phi^{(r)}\right] = \frac{1}{L}\sum_{l=1}^{L} \ln K(\gamma^{(l)}|D)$.

3. **The M-step** (Maximization step)

Find the values

(a) $\beta^{(r+1)}$ and $\phi^{(r+1)}$ that maximizes 2 (a).

(b) $D^{(r+1)}$ that maximizes 2 (b).

If convergence is achieved, then the current values are the MLEs, otherwise increment $r = r + 1$ and go to step 1.

5. Simulation Study

The proposed method is evaluated using simulated data set. The response variable is assumed to be binary variable.

5.1. Simulation Setup

The number of subjects is fixed at 100 subjects and the time points are chosen as 7 occasions. Binary responses $y_{it}, i = 1,2, \dots, 100; t = 1,2, \dots, 7$ are generated, conditionally on the random effect γ_i, from a Bernoulli distribution with mean μ_{it}. The random intercept logit model $\text{logit}(\mu_{it}) = \beta_1 + \beta_2 Trt_i + \beta_3 Time_{it} + \gamma_i$ is used.

The parameters values are chosen as $\beta_1 = -1.5$, $\beta_2 = -0.5$ *and* $\beta_3 = -0.4$. The binary covariate *Trti* is set to be 1 for half of the subjects and 0 for the other half. The continuous covariate *Time* is independently generated from normal distribution with mean vector (0 1 2 3 6 9 12) and

standard deviation vector (0 0.1 0.2 0.3 0.5 0.6 0.8). Note that, for each subject *i*, the 1*st* value of the *Time* covariate will always be 0. The random intercepts γ_i's are obtained as $\gamma_i = 4\gamma_{st,i}$ such that $sd(\gamma_i) = 4$. Standardized random intercepts $\gamma_{st,i}$'s are generated from lognormal distribution *Ln N(0; 1)*. The lognormal density is chosen to represent a skewed distribution whose support does not cover the whole real line unlike the normal distribution.

This setup is the same as in[7] to enable us to compare our results with those in[7].

In this case the GLMM in Equation (4) is

$$\begin{gathered} Y_{it}|\gamma_i \sim \ indep\ Ber(\mu_{it}) \\ \text{logit}(\mu_{it}) = \eta_{it} = \beta_1 + \beta_2 Trt_i + \beta_3\, Time_{it} + \gamma_i \qquad (8) \\ Y_{it}|\gamma_i \sim K(\gamma_i|D). \end{gathered}$$

Some remarks are in order about this model. First, $\phi = 1$ so it drops out from the calculations. Second, we have a single random effect which is common (has the same value) for all the measurements of each subject. Thus γ_i's are *iid* from $K(\gamma_i|D)$ where the variance *D* is a scalar. Finally, the distribution *K(.)* will be the lognormal distribution.

5.2. The Logit-Lognormal Model

For this model, the probability density function of the random effect is log-normal with parameters *a* and *b*,

$$K(\gamma_i|a,b) = \frac{1}{\gamma_i b\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln\gamma_i - a}{b}\right)^2},$$

where $\gamma_i \geq 0, a \in \mathbb{R}$ and $b > 0$.

The likelihood function in Equation (5) can be written as:

$$\begin{aligned} L(\beta,a,b|y) &= \prod_{i=1}^{100}\int \Pr(Y_i = y_i|\gamma_i,\beta)\, K(\gamma_i|D)\, d\gamma_i \\ &= \prod_{i=1}^{100}\int \prod_{t=1}^{7} \Pr(Y_{it} = y_{it}|\gamma_i,\beta)\, K(\gamma_i|D)\, d\gamma_i \\ &= \prod_{i=1}^{100}\int \prod_{t=1}^{7} \left(\frac{\mu_{it}}{1-\mu_{it}}\right)^{y_{it}} (1-\mu_{it}) K(\gamma_i|D) d\gamma_i \\ &= \prod_{i=1}^{100}\int \prod_{t=1}^{7} e^{\eta_{it}y_{it}}(1+e^{\eta_{it}})^{-1} \frac{1}{\gamma_i\, b\,\sqrt{2\pi}} e^{\frac{-(\ln\gamma_i - a)^2}{2b^2}} d\gamma_i \end{aligned}$$

Hence, the log-likelihood is given as:

$$\begin{aligned} l(\beta,a,b|y) = -\frac{1}{2}\ln 2\pi b^2 \sum_{i=1}^{100}\int \Bigg\{ &\sum_{t=1}^{7} \eta_{it}y_{it} \\ &- \sum_{t=1}^{7} \ln(1+e^{\eta_{it}}) \\ &- \ln\gamma_i - \frac{1}{2}\left(\frac{\ln\gamma_i - a}{b}\right)^2 \Bigg\} d\gamma_i \end{aligned}$$

We apply the MCEM algorithm introduced in Section 4.2. Starting from initial values $\beta^{(0)} = \left(\beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}\right)^T, a^{(0)}$ and $b^{(0)}$ at iteration (r+1), the algorithm proceeds as:

1. For each subject i, generate L values $\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(L)}\}$ from $\ln\mathcal{N}(a^{(r)}, b^{(r)})$ using the Metropolis algorithm with probability of acceptance as in Equation (7) given by

$$\mathcal{A}(\gamma_i, \gamma_i^*) = \min\left(1, \frac{e^{\gamma_i^*}\sum_t y_{it}\prod_t\left[1+e^{\eta_{it}^*}\right]^{-1}}{e^{\gamma_i}\sum_t y_{it}\prod_t[1+e^{\eta_{it}}]^{-1}}\right),$$

where $\eta_{it}^* = \beta_1^{(r)} + \beta_2^{(r)}Trt_i + \beta_3^{(r)}Time_{it} + \gamma_i^*$ and $\eta_{it} = \beta_1^{(r)} + \beta_2^{(r)}Trt_i + \beta_3^{(r)}Time_{it} + \gamma_i$

2. The E-Step

Calculate the expectations as Monte Carlo estimates

(a) $E\left[\ln f(y|\gamma,\beta)|y,\beta^{(r)}\right] = \frac{1}{L}\sum_{l=1}^{L}\ln f(y|\gamma^{(l)},\beta)$

$$= \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{100}\sum_{t=1}^{7}\left\{\eta_{it}^{(l)}y_{it} - \ln\left[1+e^{\eta_{it}^{(l)}}\right]\right\},$$

where

$$\eta_{it}^{(l)} = \beta_1 + \beta_2 Trt_i + \beta_3 Time_{it} + \gamma_i^{(l)}$$

(b) $E[\ln K(\gamma|a,b)|y,\beta^{(r)}] = \frac{1}{L}\sum_{l=1}^{L}\ln K(\gamma^{(l)}|a,b)$

$$\simeq -100\ln b - \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{100}\left[\ln\gamma_i^{(l)} + \frac{1}{2}\left(\frac{\ln\gamma_i^{(l)}}{b}\right)^2\right].$$

Note that $\beta^{(r)}$ do not show explicitly in this step but they already used in Step 1 to generate the L values of γ_i's.

3. The M-step

Find the values

(a) $\beta^{(r+1)}$ that maximizes 2 (a).

(b)$a^{(r+1)}$ and $b^{(r+1)}$ that maximizes 2 (b) to use them for the calculation of the standard deviation $D^{(r+1)}$.

If convergence is achieved, then the current values are the MLEs, otherwise increment r = r + 1 and return to step 1.

5.3. Simulation Results

The algorithm described above is implemented using the MATLAB package. Fifty data sets were generated according to the setup in Section 5.1. The random effect was simulated from a lognormal distribution. Then each data set is analysed using the Metropolis EM (MEM) algorithm twice. First, under the normality assumption of the random effects. Second, assuming that the random effects belong to the lognormal distribution. Summaries of the results are given in Tables 2 and 3. Note that Bias, Std. Dev. and MSE are the average bias, standard deviation and mean squared error of the estimates respectively.

Table 2. The Metropolis EM estimates of 50 datasets

Normal Distribution	β_1	β_2	β_3	$sd(\gamma)$
Bias	1.2673	-0.0048	0.4048	0.2728
Std. Dev.	0.4952	0.4138	0.0315	0.3031
MSE	1.8463	0.1678	0.1649	0.1644

Table 3. The Metropolis EM estimates of 50 datasets

Lognormal Distribution	β_1	β_2	β_3	$sd(\gamma)$
Bias	-0.2994	0.1213	0.4028	-1.2281
Std. Dev.	0.2992	0.3285	0.0209	1.1879
MSE	0.1774	0.1205	0.1627	2.8910

In general, we get better results for the fixed effects when the lognormal density is assumed for the random effect. The parameters' estimates are less variable and the values of the MSE are smaller. The lognormal MEM estimate for β_2 is more biased but this is compensated with smaller variance resulting in a lower value of the MSE. On the other hand, the estimate for the standard deviation of the random effect is better for the normal model than the lognormal model.

It is clear that there is only a small improvement concerning the estimate of β_3 when using the logormal MEM. The bias for β_3 is large when using either, normal or lognormal MEM, it is 100% of the true value of the parameter. Therefore, we turn to a modified MEM (MECM) to improve the results. The modification is in the M-step where the parameter vector β is divided into two subsets; (β_1, β_2) and (β_3). The maximization of the Monte Carlo expectations calculated in the preceding E-step is replaced by a conditional maximization. In other words, we first calculate (β_1, β_2) that maximize the expectation while β_3 is held fixed. Next, we substitute the maximization arguments for β_1 and β_2 in the expectation function and find the value of β_3 that maximizes the function. We repeat this conditional maximization twice. The results obtained for 50 data sets are summarized in Table 4.

Table 4. The lognormal MECM estimates of 50 datasets

	Lognormal Distribution			
	β_1	β_2	β_3	$sd(\gamma)$
Bias	0.2674	0.2674	0.1829	0.5581
Std. Dev.	0.4465	0.2768	0.1299	1.4588
MSE	0.2668	0.0751	0.0500	2.3971

Compared to lognormal MEM (Table 3), results are improved when the modified MEM (MECM) is used. We got smaller MSE for all estimates except for β_1 and the bias is considerably lower for β_2, β_3 and the standard deviation of the random effect $sd(\gamma)$. On the other hand, the estimates for β_1, β_3 and $sd(\gamma)$ become more variable.

Table 5. Lognormal MECM and normal GLMM for 100 data sets

	Lognormal MECM			
	β_1	β_2	β_3	$sd(\gamma)$
Bias	**0.0396**	**0.0055**	0.0731	0.7995
Std. Dev.	**0.4557**	**0.2941**	0.1558	1.6482
MSE	**0.2071**	**0.0857**	0.0294	3.3288
	Normal GLMM			
Bias	-2.2325	-0.1686	**0.0271**	**0.7975**
Std. Dev.	1.1023	1.3581	**0.0696**	**1.1778**
MSE	6.1991	1.8729	**0.0056**	**2.0233**

We are going to compare our results (the MECM method) with those given in[7]. Table 5 shows the results for the lognormal MECM algorithm of 100 data sets and also the results for the normal GLMM given in[7]. The smaller values, when comparing the two models, are written in bold. A significant improvement can be seen when the lognormal MECM algorithm is used to calculate the estimates for fixed effect β_1 and β_2. Results for β_3 are good for both lognormal MECM and normal GLMM models, however, the latter gives better values. Finally, MECM algorithm produces more variable estimates for $sd(\gamma)'$s resulting in a higher value for MSE.

6. Conclusions

In this paper, we developed a Monte Carlo EM algorithm to estimate regression parameters for a logit model with lognormal random effects. The proposed method was applied to simulated binary data. A modified M-step was used to improve the results but a trade off between small values for bias and small variability must be made. In general, the obtained results are acceptable when comparing the MEM estimates to those calculated using the normal GLMM.

Further work is to apply the proposed method to larger data sets. We can develop the MEM to logit model with different distribution for the random effect, namely, gamma distribution which is a natural conjugate for the binary data. This work is under investigation now.

REFERENCES

[1] Annis, D. H., "A note on quasi-likelihood for exponential families", Statistics and Probability Letters, 77, 431-437, 2006.

[2] Dempster, A., Laird, N., and Rubin, D., "Maximum Likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society., Series B, 39, 1-38, 1977.

[3] Diggle, P. J., Liang, K-Y., and Zeger, S. L., "Analysis of Longitudinal Data", Clarendon Press, Oxford, UK, 1994.

[4] Dobson, A. J., "An introduction to generalized linear models", Chapman and Hall, London, UK, 1990.

[5] Dunlop, D. D., "Regression for longitudinal data: a bridge from least squares regression", The American Statistician, 48, 299-303, 1994.

[6] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., "Introducing Markov Chain Monte Carlo", Chapter 1, In : Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.), Markov Chain Monte Carlo in practice, Chapman and Hall, London, 1995.

[7] Komarek, A., and Lesaffre, E., "Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution", Computational Statistics and Data Analysis, 52, 3441-3458, 2008.

[8] Laird, N. M., and Ware, J. H., "Random effects models for longitudinal data", Biometrics, 38, 963-974, 1982.

[9] Lee, Y., Nelder, J. A., and Pawitan, Y., "Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood, Chapman and Hall, London, UK, 2006.

[10] Li, E., Zhang, D., and Davidian, M. (2004), "Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements", Biometrics, 60, 1-7, 2004.

[11] Liang, K-Y., and Zeger, S. L., "Longitudinal data analysis using generalized linear models", Biometrika, 73, 13-22, 1986.

[12] McCullagh, P., "Quasi-Likelihood functions", The Annals of Statistics, 11, 59-67, 1983.

[13] McCullagh, P., and Nelder, J. A., "Generalized linear models", 2nd edition, Chapman and Hall, London, UK, 1989.

[14] McCulloch, C. E., "Generalized Linear Mixed Models", NSF-CBMS Regional Conference Series in Probability and Statistics, 7, Institute of Mathematical Statistics, 2003.

[15] McCulloch, C. E., and Searle, S. R., "Generalized, Linear, and Mixed Models", John Wiley & Sons, Inc., New York, USA, 2001.

[16] McGilchrist, C. A., "Estimation in generalized mixed models", Journal of the Royal Statistical Society, Series B, 56, 61-69, 1994.

[17] McLachlan, J., and Krishnan, T., "The EM algorithm and extensions", Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, USA, 1997.

[18] Nelder, J., and Wedderburn, R., "Generalized Linear Models", Journal of the Royal Statistical Society A, 135, 370 – 384, 1972.

[19] Proust, C., and Jacqmin-Gadda, H., "Estimation of linear mixed models with a mixture of distribution for the random effects", Computer Methods and Programs in Biomedicine, 78, 165-173, 2005.

[20] Robert, C., "Methodes de Monte Carlo par Chaines de Markov, Economica, Paris, 1996.

[21] Roberts, G. O., "Markov Chain concepts related to sampling algorithm", Chapter 3, In : Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.), Markov Chain Monte Carlo in practice, Chapman and Hall, London, UK, 1996.

[22] Steele, B., "A modified EM algorithm for estimation in generalized mixed models", Biometrics, 52, 1295-1310, 1996.

[23] Tutz, G., and Kauermann, G., "Generalized linear random effects models with varying coefficients", Computational Statistics and Data Analysis, 43, 13-28, 2003.

[24] Wedderburn, R. W., "Quasi-likelihood functions, generalized linear models and the Gauss-Newton method", Biometrika, 61, 439-447, 1974.

[25] Zeger, S. L., and Liang, K-Y., "Longitudinal data analysis for discrete and continuous outcomes", Biometrics, 42, 121-130, 1986.

[26] Zeger, S. L., Liang, K-Y., and Albert, P. S., "Models for longitudinal data: a generalized estimating equation approach", Biometrics, 44, 1049 – 1060, 1988.

[27] Zeger, S. L., Liang, K-Y., and Self, S. G., "The analysis of binary longitudinal data with time-independent covariates", Biometrika, 72, 31-38, 1985.