پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

Asymptotic Analysis of Numerically Estimated Likelihood Estimators within the Context of Error-Prone Regression Models

Maryam Ahangari*[1], Mousa Golalizadeh[1], Zahra Rezaei Ghahroodi[2]

[1]Department of Statistics, Tarbiat Modares University

[2]School of Mathematics, Statistics and Computer Science, University of Tehran

Abstract

Generalized linear mixed models (GLMMs) represent prevalent methodologies for the examination of multilevel datasets, employing the maximum likelihood framework for the estimation of model parameters. In the context of nonlinear regression models pertaining to the outcome variable, it is imperative to address classification errors associated with responses as well as measurement errors present in covariates, inorder to ensure the precision of inferences, which consequently renders the likelihood function analytically intractable. In such scenarios, a traditional yet adaptable methodology to tackle numerical integration challenges, is the Monte Carlo simulation; nonetheless, simulation analyses indicate that the utilization of alternative approximation techniques, such as the multivariate Gauss-Hermite quadrature method, demonstrates significantly superior performance in practical applications. This manuscript delineates the general conditions requisite for the Gauss-Hermite approximation technique, aimed at guaranteeing both consistency and asymptotic normality. Furthermore, a simulation study is included to exemplify this point.

Keywords: Multivariate Gauss-Hermite quadrature, Measurement error, Mislassification, Consistency, Asymptotic normality.
Mathematics Subject Classification (2020): 60F99.

1 Introduction

The application of multilevel models is to form a pattern for the relation between the response variable and a set of covariates by incorporating the correlation between repeated measurements. When the response variable is non-normally distributed, generalized linear mixed models (GLMMs) are extensively applied, considering a nonlinear link between the mean of the response variable with a set of predictors. In many situations, some of the predictors of GLMMs may not be measured preciesly, so measurement error in continuous covariates is

*Corresponding author, maryam.ahangari@gmail.com

a challenging problem in the analysis of multilevel-structured data with discrete responses. General strategies of handling covariate mismeasurement effects in estimating the parameters corresponding to correlated data is the Monte Carlo scheme. Recently, Xie and *et al.* (2017) have shown that the Monte Carlo Newton-Raphson (MCNR) method gives accurate estimates in the logistic regression model for analyzing longitudinal biomarker data, accounting for left-censoring and covariate measurement error.

As a distinct source of error, the misclassification of categorical response variables within the context of regression analysis represents a significant consideration in evaluating the validity of parameter estimation as well as in facilitating subsequent statistical inferences. Consequently, the rectification of misclassification has emerged as a focal point for researchers aiming to achieve precise estimates alongside valid confidence intervals. For further clarification, the research conducted by Buonaccorsi (2010) is recommended. In the context of response misclassification, Neuhaus (2002) assessed the extent of asymptotic bias inherent in the estimation of logistic regression parameters when the response is subject to classification error within a longitudinal binary data framework. Paulino and *et al.* (2003) have additionally introduced a Bayesian paradigm for analyzing binary responses that are affected by classification error, employing a random effects logistic regression model. Furthermore, Tang and *et al.* (2015) concentrated on longitudinal validation data-driven studies that involve repeatedly measured responses characterized by measurement error. They assumed a differential misclassification mechanism and employed the maximum likelihood approach to correct for misclassification in the binary response.

Measurement error in covariates together with misclassification in the outcomes occur frequently in epidemiological studies, and failure to account for these two major sources of error, results in potentially biased and inaccurate estimates. Roy (2012) has conducted a comprehensive examination of these dual sources of errors within the framework of longitudinal data, aiming to rectify both response classification inaccuracies and covariate measurement discrepancies, through the application of the Monte Carlo Markov Chain (MCMC) computational methodology. Yi (2017) has investigated methodologies for concurrently addressing errors in both response variables and covariates, pertaining to univariate as well as correlated datasets.

The structure of this article is as follows: In Section 2, we describe generalized linear mixed models involving binary correlated response components in conjunction with non-differential misclassified responses and covariate measurement error when covariates are subject to measurement error. We also discuss the likelihood function and the application of the MGHQ method in scenarios where precise measurements of certain covariates are unavailable. In Section 3, we focus on the conditions required to ensure the consistency and asymptotic normality of estimators based on such approximations. Finally, in Section 4 we apply the theory to a simulation study based on a random effects logistic regression model subject to response misclassification and covariate measurement error.

2 The Generalized Linear Mixed Misclassification and Measurement error Model, Basics and Notation

Suppose $Y_i = (Y_{i1}, \cdots, Y_{in_i})^T$ denotes the observed outcomes for the ith subject, $1 \leq i \leq m$, where m is the total number of independent individuals, and n_i is the number of observations for individual i. Let Y_i be a misclassification-prone response variable that follows a generalized linear mixed model with random intercept for each individual. The vector of model covariates $X_i = (X_{i1}^T, \cdots, X_{in_i}^T)^T$ and $Z_i = (Z_{i1}^T, \cdots, Z_{in_i}^T)^T$ are assumed to be time-varying. Let Y_{ij} be the response variable for subject i at the jth occasion measurement ($j = 1, \cdots, n_i$), X_{ij} as the vector of true covariates which are error-prone and unobserved and hence latent, and Z_{ij} be the vector of error-free covariate. Furthermore, let $X_{ij}^\star$ and $Y_{ij}^\star$ be surrogate measurements of X_{ij} and Y_{ij}, respectively, that are observed in the absence of X_{ij} and Y_{ij}. Note that the

notations used here are similar with those utilized in the literature. See, e.g., Ahangari and *et al.* (2019)

2.1 The Outcome Process, Measurement Error Models and Classification Error Probabilities

Let Y_{ij} be a binary response variable for subject i at occasion j. Assume that the binary outcome is associated with the collection of covariates via the random effects logistic regression framework as

$$logit[P(Y_{ij}=1|x_{ij},z_{ij},\tau_i)]=\beta_0+\beta_x x_{ij}+\beta_z z_{ij}+\tau_i, \quad i=1,\cdots,m, \quad j=1,\cdots,n_i. \tag{1}$$

In (1), $\beta=(\beta_0,\beta_x,\beta_z)^T$ is the vector of fixed parameters. We assume that the random effects τ_i, $(i=1,\cdots,m)$ are independent from each other, and also independent from the error-prone covariates X_{ij}. In this article, it is assumed that $\tau_i \sim N(0,\sigma_\tau^2)$, where σ_τ^2 is unknown and assumed to be constant for all subjects.
In (1), the covariate X_{ij} is not directly observable, and $X_{ij}^\star$ is the only available measurement in the absence of X_{ij}, so $X_{ij}^\star$ can be regarded as the proxy variable for X_{ij}. In the following, it is assumed that $X_{ij}^\star$ is a combination of X_{ij} plus some error, specifically classical measurement error, in which the true but latent variable is measured with additive error, usually assumed to have constant variance, i.e.

$$\boldsymbol{X_{ij}^\star|x_{ij}=x_{ij}+e_{ij}.} \tag{2}$$

where $\boldsymbol{e_{ij}}$ is the error term such that $E(e_{ij}|x_{ij})=0$ and $Var(e_{ij}|x_{ij})=\sigma_e^2$, when $X_{ij}^\star$ and X_{ij} are scalar. It is concluded that $E(X_{ij}^\star|x_{ij})=x_{ij}$, so $X_{ij}^\star$ is unbiased for the unobserved x_{ij}. We also consider that the measurement error in $X_{ij}^\star$ is non-differential. This means that Y_{ij} is conditionally independent of $X_{ij}^\star$, given x_{ij} and if X_{ij} is observed, the proxy variable $X_{ij}^\star$ adds nothing to the predictor of Y_{ij}, i.e.;

$$P(Y_{ij}=1|x_{ij},x_{ij}^\star,z_{ij},\tau_i)=P(Y_{ij}=1|x_{ij},z_{ij},\tau_i). \tag{3}$$

In the following, it is assumed that we have homoscedastic measurement error, i.e. the variance of W_{ij} given x_{ij} is constant.
With random covariate X_i subject to error, we assume distributional assumption on error-prone covariate X_i, i.e. structural approach in which covariate X_i has a normal distribution with mean μ_x and variance Σ_x, and is independent of τ_i. We choose a fully structural case for the mismeasured covariate with X_i assumed i.i.d with

$$X_i=(X_{i1}^T,\cdots,X_{in_i}^T)^T \sim N(\mu_x,\Sigma_x) \quad i=1,\cdots,m.$$

The measurement error in the multivariate case, also has a normal distribution with the following structure:

$$e_i \sim N(0_{n_i},\sigma_e^2 I_{n_i}).$$

Suppose that Y_{ij} is also subject to misclassification, and Y_{ij}^* is an observed value of Y_{ij}. Particularly, let

$$P(Y_{ij}^*=1|Y_{ij}=0,x_{ij},z_{ij},\tau_i)=P(Y_{ij}^*=1|Y_{ij}=0)=\gamma_{01} \tag{4}$$

and

$$P(Y_{ij}^*=0|Y_{ij}=1,x_{ij},z_{ij},\tau_i)=P(Y_{ij}^*=0|Y_{ij}=1)=\gamma_{10}, \tag{5}$$

denote the simple case of misclassification probabilities. The probability $1-\gamma_{10}$ is often called the sensitivity of the measurement Y_{ij}^*, and $1-\gamma_{01}$ is called the specificity. In (4) and (5), γ_{01} and γ_{10} are treated as unrecognized, nonnegative classification error probabilities

that are constant and independent of other covariates within the model. This declaration is generally known as 'non-differential misclassification'. According to the notations given so far, the conditional probability for the observed response given the true values of covariates and the random component is provided as follows:

$$\begin{aligned} P_{ij}^{*} &= P(Y_{ij}^{*}=1|x_{ij},z_{ij},\tau_i) \\ &= P(Y_{ij}^{*}=1|Y_{ij}=1,x_{ij},z_{ij},\tau_i)P(Y_{ij}=1|x_{ij},z_{ij},\tau_i) \\ &+P(Y_{ij}^{*}=1|Y_{ij}=0,x_{ij},z_{ij},\tau_i)P(Y_{ij}=0|x_{ij},z_{ij},\tau_i) \\ &= (1-\gamma_{10})P_{ij}+\gamma_{01}(1-P_{ij})=\gamma_{01}+(1-\gamma_{10}-\gamma_{01})P_{ij}, \end{aligned} \tag{6}$$

where the quantity P_{ij} is described as

$$P_{ij}=P(Y_{ij}=1|x_{ij},z_{ij},\tau_i)=\frac{\exp(\beta_0+\beta_x x_{ij}+\beta_z z_{ij}+\tau_i)}{1+\exp(\beta_0+\beta_x x_{ij}+\beta_z z_{ij}+\tau_i)}. \tag{7}$$

Besides, let us consider $X_{ij}^{\star}$ as a surrogate for X_{ij} with classical additive measurement error as (2), where $e_{ij}\sim N(0,\sigma_e^2)$. Then, the conditional distribution of $X_{ij}^{\star}|x_{ij}$ can be written as $X_{ij}^{\star}|x_{ij}\sim N(x_{ij},\sigma_e^2)$.

2.2 Likelihood Function Arrangements

In this Section, we provide a parametric likelihood procedure that allows for classification error in the binary responses and measurement error in the covariates in a random effects logistic regression model. According to (1) and (2), if $\theta=(\beta,\sigma_\tau^2,\sigma_e^2)^T$ is a vector of associated parameters along with the vector of misclassification rates $\gamma=(\gamma_{01},\gamma_{10})^T$, for the ith subject at occasion j, given the observed data $(y_{ij}^{*},x_{ij}^{\star},z_{ij})$ for $i=1,\cdots,m$ and $j=1,\cdots,n_i$, the log-likelihood function inspired by the observed data encompassing the measurement error distribution and the classification errors can be written as follows:

$$\ell(\theta;y^{\star},x^{\star})=\sum_{i=1}^{m}\log\int_R\int_{R^{n_i}}\prod_{j=1}^{n_i}f(y_{ij}^{\star}|x_{ij},z_{ij},\tau_i;\beta)f(x_{ij}^{\star}|x_{ij};\sigma_e^2)f(x_i;\mu_x,\sigma_{xj}^2)f(\tau_i;\sigma_\tau^2)dx_{ij}d\tau_i. \tag{8}$$

For the approximation of the integral (8), multivariate forms of the Gauss-Hermite quadrature approximation method is needed to handle both unobserved covariate x_{ij} at n_i occasions and the random effects τ_i. Due to the fact that both x_{ij} and τ_i are independent from each other, the product of their density functions can be defined as a multivariate normal distribution.
The multivariate Gauss-Hermite quadrature (MGHQ) method extends the univariate Gauss-Hermite quadrature rule by applying it to each coordinate of the integrated variables. To approximate of the integrals in (8) which are (n_i+1) dimensional, a matrix of GHQ nodes are generated for unobseved variables X_{ij} at occasion j and also for the random effects τ_i. This is done based on the mean vector and variance-covariance matrix of the multivariate normal distribution $f(x_i,\tau_i;\sigma_{xj}^2,\sigma_\tau^2)$, with respect to which the integral is calculated. Additionally, a vector of weights corresponding to this multivariate normal distribution is computed to complete the approximation. The problem is that the integrals in (8) are multidimensional, with their values depending upon both the random effects as well as the underlying latent covariate X_{ij} across n_i occasions. The estimated parameter vector $\hat{\theta}_{ML}\in\Theta\subset R^p$ is acquired by maximizing $L(\theta;y^{\star},x^{\star})$, with p as the number of parameters in the regression model. As previously noted, the integrals in (8) cannot be evaluated in a closed form. To manage unobserved covariates and random effects, multivariate Gauss-Hermite quadrature (MGHQ) approximation approach is used, by choosing sets of GHQ nodes $\hat{x}_{k_j}=(\hat{x}_{k_1 j}^{(1)},\hat{x}_{k_2 j}^{(2)},\cdots,\hat{x}_{k_{(n_i+1)}j}^{(n_i+1)})'$, $\hat{\tau}_k=(\hat{\tau}_{k_1}^{(1)},\hat{\tau}_{k_2}^{(2)},\cdots,\hat{\tau}_{k_{(n_i+1)}}^{(n_i+1)})'$ in addition with a vector of weights $w_k=(w_{k_1}^{(1)},w_{k_2}^{(2)},\cdots,w_{k_{(n_i+1)}}^{(n_i+1)})'$. In this setting, the parameter k refers to the indices $k_1,k_2,\cdots,k_{(n_i+1)}$, each index shows different sets of quadrature points assigned, in order to approximate each of (n_i+1) integrals. Consequently, according to Heiss and Winschel (2008), the MGHQ approximation to the log-likelihood function of (8) will result the following

structure:

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z) &= \sum_{i=1}^{m} \log \Bigg[\sum_{k_1=1}^{q_1} \sum_{k_2=1}^{q_2} \cdots \sum_{k_{(n_i+1)}=1}^{q_{(n_i+1)}} w_{k_1}^{(1)} w_{k_2}^{(2)} \cdots w_{k_{(n_i+1)}}^{(n_i+1)} \\ &\quad \times \prod_{j=1}^{n_i} f(y_{ij}^\star | x_{kj}, z_{ij}, \tau_k; \beta) f(x_{ij}^\star | x_{kj}; \sigma_e^2) \Bigg] \\ &= \sum_{i=1}^{m} \log \Bigg[\sum_{k_1=1}^{q_1} \sum_{k_2=1}^{q_2} \cdots \sum_{k_{(n_i+1)}=1}^{q_{(n_i+1)}} w_{k_1}^{(1)} w_{k_2}^{(2)} \cdots w_{k_{(n_i+1)}}^{(n_i+1)} \prod_{j=1}^{n_i} \phi(y_{ij}^\star, x_{kj}, z_{ij}, \tau_k) \Bigg].\end{aligned} \tag{9}$$

On the whole, this technique results the maximum approximated likelihood estimator (MALE), i.e.

$$\hat{\theta}_{MALE} = arg\ max_{\theta \in \Theta}\ \ \tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z).$$

The selection of appropriate quadrature points and weights is crucial to achieving the desired level of accuracy in numerical integration. While increasing the parameter k enhances the approximation accuracy, it also incurs a corresponding increase in computational cost. Consequently, it is essential to determine the necessary level of accuracy that guarantees the resulting estimator maintains both consistency and asymptotic normality. To address this, we introduce a link function $K : \mathbb{N} \to \mathbb{N}$ that relates the number of quadrature points k to the sample size n, in other words $k = K(n)$. In the context of Monte Carlo techniques, the likelihood function is approximated through simulation, and consistency is ensured under standard regularity conditions provided that the number of simulation draws k increases with the sample size, i.e. $k = K(n) \geq n$.

3 Asymptotic Theory Corresponding to Approximated Likelihood Function

As previously noted, when the likelihood contribution $\ell(\boldsymbol{\theta}; y^\star, x^\star, z)$ cannot be derived in closed form due to its analytical intractability, it becomes necessary to employ an approximation method. The approximated likelihood $\tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z)$ depends on an accuracy parameter k. This parameter shows the degree of approximation error as well as the interrelated computational costs. Accordingly, we define the quantity

$$\epsilon(k) = sup_{\theta \in \Theta} |\tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z) - \ell(\boldsymbol{\theta}; y^\star, x^\star, z)| + sup_{\theta \in \Theta} ||\nabla_\theta \tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z) - \nabla_\theta \ell(\boldsymbol{\theta}; y^\star, x^\star, z)||, \tag{10}$$

which measures the worst-case approximation error of the function ℓ and its gradiant $\tilde{\ell}$.
Beyond establishing the consistency of the maximum likelihood estimator, the only additional requisite for the consistency of $\hat{\theta}_{MALE}$ is the uniform convergence in probability of the surrogate log-likelihood $\tilde{\ell}(\boldsymbol{\theta})$ to the true, but analytically intractable $\ell(\boldsymbol{\theta})$. To rigorously pursue this aim, it is expedient to adopt the following notation for the partial derivative of a sufficiently smooth function g:

$$D_\alpha^{(\theta)} g(\theta) = \frac{\partial^{|\alpha|}}{\prod_{j=1}^{p} \theta_j^{\alpha_j}} g(\theta),$$

where $\alpha = (\alpha_1, \cdots, \alpha_p)$ is a multi-index indicating the order of partial derivatives for different parameters $\theta_1, \cdots, \theta_p$. In the subsequent analysis, for the purpose of establishing an upper bound on the magnitude of $sup_{\theta \in \Theta} |\tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z) - \ell(\boldsymbol{\theta}; y^\star, x^\star, z)|$ together with $sup_{\theta \in \Theta} ||\nabla_\theta \tilde{\ell}(\boldsymbol{\theta}; y^\star, x^\star, z) - \nabla_\theta \ell(\boldsymbol{\theta}; y^\star, x^\star, z)||$, we bring into action

$$\bar{\epsilon}_r(k) = max_{|\alpha|_1 \leq r, \theta \in \Theta} sup |D_\alpha^{(\theta)} \ell(\boldsymbol{\theta}; y^\star, x^\star, z) - D_\alpha^{(\theta)} \tilde{\ell}_k(\boldsymbol{\theta}; y^\star, x^\star, z)| \tag{11}$$

for $r = 0, 1, 2$; indicating all partial derivatives with respect to θ up to order r, (Griebel and *et al.*, 2019).

3.1 Consistency and Asymptotic Normality

In the ensuing Section, we shall formulate comprehensive conditions on the log-likelihood contribution (8) and the approximated function (9) to guarantee the desirable asymptotic properties of the MAL estimator. We begin by revisiting the assumptions underpinning the consistency of the maximum likelihood estimator as established by Newey and McFadden (1994).

Lemma 3.1. *Assume that (i) For all $\theta \neq \theta_0$, we have $f(y;\boldsymbol{\theta}) \neq f(y;\boldsymbol{\theta_0})$; (ii) $\theta_0 \in \Theta$ and Θ is compact; (iii) $\log f(y;\boldsymbol{\theta})$ is continuous at each $\theta \in \Theta$ with probability one; (iv) $\mathbb{E}_y[sup_{\theta\in\Theta}|\log f(y;\boldsymbol{\theta})|] < \infty$. Then, $plim_{n\to\infty}\hat{\theta}_{ML} = \theta_0$ (consistency).*

We next invoke the asymptotic normality of the ML estimator.

Lemma 3.2. *Assume that (i) the assumptions of Lemma 3.1 hold and that (ii) $\theta_0 \in \Theta$; (iii) $f(y;\boldsymbol{\theta})$ is twice continously differentiable and $f(y;\boldsymbol{\theta}) > 0$ in a neighborhood $\mathcal{N}$ of θ_0; (iv) $\int sup_{\theta\in\mathcal{N}}||\nabla_\theta f(y;\boldsymbol{\theta})||dy < \infty$ and $\int sup_{\theta\in\mathcal{N}}||\nabla_{\theta\theta} f(y;\boldsymbol{\theta})||dy < \infty$; (v) The function defined by $\mathcal{L} = \mathbb{E}_y[\nabla_\theta f(y;\boldsymbol{\theta_0})(\nabla_\theta f(y;\boldsymbol{\theta_0}))^T]$ exists and is nonsingular, (vi) The last assumption is $\mathbb{E}_y[sup_{\theta\in\mathcal{N}}||\nabla_\theta \log f(y;\theta_0)||] < \infty$. Then $\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, \mathcal{L}^{-1})$ (asymptotic normality).*

According to Lemma 3.1 for an assumed function $f(y;\theta)$, in order to handle with consistency of MALE, we use the specific setting where the log-likelihood has the form (8) and the MGHQ-approximated function follows the (9) contribution. For more details, see Griebel and *et al.* (2019).

Corollary 3.3. *Assume that (i) there exists $\bar{\delta} > 0$ such that for all $\theta \in \Theta$ it holds that $l(\boldsymbol{\theta}; y^\star, x^\star, z) > \bar{\delta} > 0$; (ii) the function $\phi(y^\star_{ij}, x_{kj}, z_{ij}, \tau_k)$ is continous in $\theta \in \Theta$; (iii) With accuracy parameter k, $plim_{k\to\infty}\bar{\epsilon}_0(k) = 0$; (iv) $K(n)$ is monotonically increasing in n. Then, $\hat{\theta}_{MAL}$ is a consistent estimator of θ_0, i.e.*

$$\hat{\theta}_{MAL} \xrightarrow{p} \theta_0. \tag{12}$$

To characterize the asymptotic distribution of the MALE, it is essential that the worst-case approximation error (11) not only converges to zero but does so at a sufficiently fast rate.

Corollary 3.4. *Assume that the assumptions of Lemma 3.2 and Corollary 3.3 hold. Also assume that (i) for all $(y^\star, x^\star, z)$ it holds that $\phi(y^\star_{ij}, x_{kj}, z_{ij}, \tau_k)$ is twice continously differentiable; (ii) for all $(y^\star, x^\star, z)$ and $\theta \in \Theta$ all terms in (8) remain bounded. (iv) $plim_{k\to\infty}\bar{\epsilon}_2(k) = 0$; (v) $R(n)$ rises fast enough with n to ensure $\sqrt{n}\bar{\epsilon}_1(K(n)) \xrightarrow{p} 0$. Then*

$$\sqrt{n}(\hat{\theta}_{MAL} - \theta_0) \xrightarrow{d} \boldsymbol{N}(0, \mathcal{L}^{-1})$$

Remark 3.5. *It is important to note that the positivity of the integration weights is not assumed at any point. Additionally, it should be observed that in practical applications, both n and $k = K(n)$ must be finite quantities, with $K(n)$ defined such that $K(n) \geq k_0(\delta)$, for the reason that otherwise $\bar{\ell} > \delta$ might not hold true.*

4 Simulations and Inferences

In this Section, we undertake simulation studies to evaluate and compare the efficiency of the Multivariate (Multidimensional) Gauss-Hermite Quadrature (MGHQ) approximation method in analyzing multilevel data subject to measurement error in the covariates and misclassification in the binary response., i.e. the hybrid method with Monte Carlo Expectation Maximization (MCEM) algorithm, as a maximum simulated likelihood (MSL) method. We use (1), i.e. logistic random effects model to simulate data. We assume $(\beta_0, \beta_x, \beta_z)$ are the intercept and covariates's fixed effects coefficients, for $i = 1, \cdots, m$ individuals and $j = 1, \cdots, n_i$ occasions. We assume that z_{ij} is an error-free or exactly measured covariate generated as a normally distributed variable following $N(0, 4^2)$ distribution, which is treated fixed during

the simulation study. The measurement error model that we have considered is the classical additive model. In this case, we generate a surrogate variable W_{ij} for the error-prone covariate X_{ij}, as in (2), where e_{ij} is the measurement error variable independently and identically distributed following $N(0, \sigma_e^2)$, but we will change the magnitude of σ_e^2 (measurement error in X_{ij}) along with different scopes of misclassification rates during the simulation study. The response variables are generated from Bernoulli distribution. We generate 2000 different data sets, each simulated data contains $m = 500$ individuals with $n_i = 3$ replicate measurements at each occasion.

4.1 Simulation Results

To assess the performance of the MGHQ and MCEM approaches, we computed simulated results in terms of the mean absolute bias, root of mean squared error (RMSE) across simulation runs, and coverage probability. The notation MGHQk is used to denote the MGHQ scenario with k quadrature points. To mitigate approximation error, the number of quadrature points was progressively increased from 2 to 20. Table 1 presents the estimated parameter values alongside the RMSEs, which serve as a standard metric for illustrating the convergence behavior of the two methods under consideration. The results in Table 1 indicate that the MGHQ approach yields accurate parameter estimates even with a small number of quadrature points, with additional points producing only marginal changes. The MCEM algorithm achieved convergence after 500 iterations. Based on the simulation outcomes, the MGHQ method exhibits lower bias in parameter estimation, reduced RMSEs, and superior coverage probabilities for nominal confidence intervals compared to the MCEM approach.

It is important to recognize that the convergence rate of an approximation method reflects the necessary growth rate of the link function $K(n)$ to ensure asymptotic normality. Therefore, we have conducted an analysis on the convergence of $\sqrt{n}\epsilon(K(n))$ for various link functions. Based on our simulation results (not presented herein), it was observed that the MGHQ method exhibits superior performance relative to the MCEM approach when utilizing a logarithmic link function.

Furthermore, the magnitude of the approximation error is intrinsically dependent on the choice of the quadrature parameter, k. This error component inevitably contributes to an increase in the variance of the estimator. In practical applications, a pragmatic strategy to mitigate this error is to progressively increase the value of k. Additionally, enlarging the sample size provides an effective means to reduce sampling variability and attain more accurate approximation results.

Table 1: True values, Parameter estimations and RMSEs with small degree of measurement error variance for two simulated and approximated methods (RMSEs in parentheses)

	Parameters						
Method	β_0	β_x	β_z	σ_τ^2	σ_e^2	γ_{01}	γ_{10}
True Value	0.02	0.6	0.5	2.4	0.5	0.05	0.05
MCEM	0.028 (0.060)	0.687 (0.086)	0.5069 (0.062)	2.707 (0.505)	0.665 (0.166)	0.040 (0.040)	0.039 (0.041)
MGHQ2	0.020 (0.055)	0.587 (0.084)	0.486 (0.050)	2.257 (0.480)	0.655 (0.158)	0.046 (0.023)	0.042 (0.022)
MGHQ3	0.017 (0.002)	0.591 (0.075)	0.489 (0.054)	2.309 (0.512)	0.651 (0.155)	0.049 (0.026)	0.042 (0.021)
MGHQ4	0.025 (0.010)	0.577 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.039 (0.026)	0.043 (0.021)
MGHQ5	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ6	0.020 (0.010)	0.571 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ7	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ8	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ9	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ10	0.020 (0.010)	0.507 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ15	0.020 (0.010)	0.500 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ20	0.020 (0.010)	0.500 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)

Discussion and Results

This paper investigates maximum approximated likelihood (MAL) estimators that extend the framework of Monte Carlo simulated likelihood (MSL) estimators for likelihood inference in generalized linear mixed models with repeatedly measured covariates subject to measurement error and misclassified binary responses. A principal advantage of MSL lies in the asymptotic properties guaranteed under broad regularity conditions due to its reliance on Monte Carlo simulation methods. However, the substantial computational burden required to attain sufficient approximation accuracy often renders this approach impractical. Consequently, more precise numerical approximation techniques, such as Gaussian quadrature, have become standard practice. The MAL estimators developed herein demonstrate robustness in handling measurement error and misclassification, which are common issues in real-world multilevel data. This robustness enhances the practical applicability of the generalized linear mixed model framework in complex settings. We have delineated a set of conditions on the approximation method that ensure the consistency and asymptotic efficiency of MAL estimators. Overall, simulation results demonstrate that the Multivariate Gauss-Hermite Quadrature (MGHQ) method yields superior parameter estimates compared to the simulation-based MCEM approach. Moreover, the MGHQ method significantly reduces computational time while maintaining or improving estimation accuracy, thereby making it a more viable choice for large-scale or high-dimensional data analyses.

References

Ahangari, M., Golalizadeh, M., and Rezaei Ghahroodi, Z. (2019), Likelihood inference in the random effects logistic regression with response misclassification and Covariate subject to measurement error, *Journal of Statistical Research of Iran*, **16(1)**, 255-286.

Buonaccorsi, J.P. (2010), *Measurement Error, Models, Methods and Applications*, New York: CRC Press.

Griebel M., Heiss F., Oettershagen, J. and Weiser, C. (2019), *Maximum Approximated Likelihood Estimation*, Bonn, Institute for Numerical Simulation.

Heiss F. and Winschel V. (2008), Likelihood approximation by numerical integration on sparse grids, *Journal of Econometrics,* **144**, 62-80.

Neuhaus, J.M. (2002), Analysis of clustered and longitudinal binary data subject to response misclassification, *Biometrics*, **58(3)**, 675-683.

Newey W. and McFadden D. (1994), *Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics*, Amsterdam, Elsevier.

Paulino, C. D., Soares, P. and Neuhaus, J. (2003), Binomial regression with misclassification, *Biometrics*, **59(3)**, 670-675.

Roy, S. (2012), Accounting for response misclassification and covariate measurement error using a random effect logit model, *Communications in Statistics-Simulation and Computation*, **41(9)**, 1623-1636.

Tang, L., Lyles, R.H., King, C.C., Hogan, J.W. and Lo, Y. (2015), Regression analysis for differentially misclassified correlated binary outcomes, *Journal of the Royal Statistical Society. Series C, Applied Statistics*, **64(3)**, 433-449.

Xie, X., Xue, X. and Strickler, H.D. (2017), Generalized linear mixed model for binary outcomes when covariates are subject to measurement errors and detection limits, *Statistics in Medicine*, **37**, 119-136.

Yi, G.Y. (2017), *Statistical Analysis with Measurement Error or Misclassification.* New York: Springer.