Effects on Regression Estimates under Misspecified Generalized Linear Mixed Models for Counts Data

Kwang Mo Jeong[1]

[1]Department of Statistics, Pusan National University

Abstract

The generalized linear mixed model(GLMM) is widely used in fitting categorical responses of clustered data. In the numerical approximation of likelihood function the normality is assumed for the random effects distribution; subsequently, the commercial statistical packages also routinely fit GLMM under this normality assumption. We may also encounter departures from the distributional assumption on the response variable. It would be interesting to investigate the impact on the estimates of parameters under misspecification of distributions; however, there has been limited researche on these topics. We study the sensitivity or robustness of the maximum likelihood estimators(MLEs) of GLMM for counts data when the true underlying distribution is normal, gamma, exponential, and a mixture of two normal distributions. We also consider the effects on the MLEs when we fit Poisson-normal GLMM whereas the outcomes are generated from the negative binomial distribution with overdispersion. Through a small scale Monte Carlo study we check the empirical coverage probabilities of parameters and biases of MLEs of GLMM.

Keywords: Generalized linear mixed model, random effects distribution, overdispersion, negative binomial responses, Kullback-Leibler information, coverage probability, bias.

1. Introduction

Generalized linear mixed models are a very flexible approach for the clustered or longitudinal outcomes of categorical responses. GLMM belongs to a class of hierarchical models that are particularly useful to structure multiple sources of variation associated with covariates, and variation attributable to unmeasured factors of subject effects. There have been significant computational advances in the routine fitting of various kinds of GLMMs via commercial packages such as SAS and R.

The estimation of regression parameters is based on the maximum likelihood theory and assumes that the underlying probability model is correctly specified. However, there has been limited works

[1]Professor, Department of Statistics, Pusan National University, Jangjeon-Dong, Kumjung-Gu, Pusan 609-735, Korea. E-mail: kmjung@pusan.ac.kr

on the sensitivity of estimates of parameters when the underlying assumptions of GLMM are violated. Some research showed that the results obtained from these models are not always robust against departures from the assumptions on which these models are based. In GLMM for counts data we usually assume a Poisson distribution for outcomes and also assume normal distribution for the random effects associated with subjects. We sometimes encounter departures from these distributional assumptions. Some researchers have studied the effects on the regression coefficients when the underlying normal assumption of random effects is violated. Neuhaus *et al.* (1992) studied the magnitude of bias of logistic-normal GLMM with misspecified mixing distributions, and found that the magnitude appears to be small even when the regression coefficients are asymptotically biased. However, Molenberghs *et al.* (1998) discussed the impact of misspecification in the respect of test performance. Recently, Litiere *et al.* (2007) studied the impact of misspecifying the random effects distribution on Type I and Type II errors in the logistic-normal random intercepts models. Similarly, Litiere *et al.* (2008) investigated the impact of misspecified random effect distribution in the GLMM of binary responses. Furthermore, Alonso *et al.* (2010) suggested a testing procedure for misspecification in logistic-normal GLMM.

In this paper we investigate the effects of misspecification on the estimates of regression coefficients in GLMM of count responses. To check the sensitivity under departures from normality we consider four kinds of random effect distributions such as normal, gamma, mixture of normals, and exponential distributions. As a second point of the paper we are also interested in the effects of misspecified distribution as Poisson whereas the true distribution is negative binomial with overdispersion. This paper consists of following sections. In Section 2 we briefly review GLMM for count responses, and explain the asymptotic normality of MLEs under misspecified models in relation to the Kullback-Leibler information criterion. In Section 3, through a Monte Carlo study we investigate the impact on estimates of parameters in terms of empirical coverage probability and bias. Finally we summarize the paper with comments on further research.

2. Generalized Linear Mixed Models for Count Responses

2.1. Poisson-Normal GLMM

Let $\{y_{ij}, \mathbf{x}_{ij}\}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, T_i$ be a clustered data consisting of n subjects having T_i elementary units, where y_{ij} denotes count responses and $\mathbf{x}_{ij} = (1, x_{1ij}, \ldots, x_{pij})'$ is $(p + 1)$-dimensional covariate values including constant. Given random subject effects u_i, the random intercepts GLMM for count responses is represented in terms of log link and mean $\mu_{ij} = E(y_{ij}|u_i, \mathbf{x}_{ij})$ as

$$\log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$. We simply let $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}$ be the fixed part of linear predictor in (2.1). We usually assume that count responses y_{ij} are Poisson distributed and the random effects u_i are independent and $N(0, \sigma^2)$, and in this case we call it the Poisson-normal GLMM.

The likelihood function $L(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma)'$ is obtained by integrating out with respect to the density of random effects.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(y_i|u_i)f(u_i)du_i, \tag{2.2}$$

where $f(\mathbf{y}_i|u_i)$ is the Poisson density of $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT_i})'$ with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT_i})'$ characterized by (2.1) and $f(u_i)$ is the density of random effects u_i. To find the MLE of $\boldsymbol{\theta}$ we need a numerical approximation of (2.2) and then a maximization step. An adaptive Gaussian quadrature or Laplacian method are commonly applied to approximate the marginal integral numerically. The random effects u_i are usually predicted by empirical Bayes method. Under some regularity conditions, the MLE $\hat{\boldsymbol{\theta}}$ satisfies $\sqrt{n}$-consistency and the asymptotic normality. We may refer to McCullagh and Nelder (1989), Breslow and Clayton (1993), and Agresti (2002) for the general discussions on GLMM. There are many available software fitting GLMM, for example, *NLMIXED* in SAS and *glmmML* in R can be used to fit Poisson-normal GLMM.

2.2. Maximum likelihood estimation under misspecification

Let $F(y, \theta)$ be an assumed probability distribution having a density $f(y, \theta)$, and also let $g(y)$ be a true probability density with its probability distribution $G(\cdot)$. The Kullback-Leibler information criterion is defined as

$$I(g; f, \theta) = E\left(\log\left[\frac{g(y)}{f(y, \theta)}\right]\right), \tag{2.3}$$

where $E(\cdot)$ is taken with respect to the true distribution $G(\cdot)$. Intuitively, $I(g; f, \theta)$ measures the ignorance about the true model. The $I(g; f, \theta)$ has the important property of $I(g; f, \theta) \geq 0$ for all $\theta \in \Theta$, where Θ is a subset of q-dimensional Euclidean space. The probability density under GLMM is determined from the conditional density of y_{ij} given random effects u_i and the random effects distribution itself. Hereafter the probability distributions $F(y, \theta)$ or $G(y)$ is associated with the marginal distribution of y_{ij} mentioned above. They may be misspecified in the respect of either the conditional distribution of y_{ij} or the random effects distribution.

According to White (1982) the MLEs $\hat{\theta}$ under $f(y, \theta)$ has the asymptotic properties: under mild regularity conditions, $\hat{\theta}_n \to \theta_*$ almost surely as $n \to \infty$, where θ_* denotes the value minimizing $I(g; f, \theta)$ of (2.3). Furthermore, $\hat{\theta}_n$ has the asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta_*) \to N(0, V_{\theta_*}), \tag{2.4}$$

where V_{θ_*} is a certain covariance matrix depending on θ_*. In particular, when $G(y) = F(y, \theta_0)$ for some $\theta_0 \in \Theta$, that is, if the probability model is correctly specified for θ_0, then the MLE is consistent estimator of θ_0. As commented by White (1982) we should be cautious about the fact that the correct specification of the probability model is only a sufficient condition but by no means necessary. As discussed by Heagerty and Kurland (2001) the θ_* can equivalently be obtained as a solution of

$$\lim_n E\left[\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(\mathbf{y}_i, \theta)|_{\theta_*}\right] = 0. \tag{2.5}$$

Our interest is whether θ_* still equals θ_0 when the underlying distributions of the random effects or the outcomes are misspecified, and, if not, it needs to be studied on the magnitude of the difference between θ_* and θ_0. In a later chapter we are to investigate the impact on MLEs in the respects of empirical coverage probabilities and biases.

Table 3.1. Coverage probabilities of correctly specified Poisson-normal GLMM when random effects distribution is $u_i \sim N(0, \sigma^2)$

n	T	σ	$1-\alpha$	β_0	β_1	β_2	σ
			0.90	0.895	0.900	0.891	0.854
		0.5	0.95	0.951	0.950	0.938	0.912
			0.99	0.984	0.992	0.985	0.969
			0.90	0.891	0.896	0.884	0.843
	3	1.0	0.95	0.928	0.944	0.934	0.904
			0.99	0.984	0.992	0.983	0.958
			0.90	0.884	0.914	0.881	0.771
		2.0	0.95	0.937	0.956	0.940	0.850
			0.99	0.983	0.996	0.986	0.936
50			0.90	0.900	0.902	0.890	0.865
		0.5	0.95	0.947	0.952	0.941	0.915
			0.99	0.983	0.995	0.988	0.975
			0.90	0.876	0.892	0.886	0.837
	5	1.0	0.95	0.935	0.938	0.944	0.907
			0.99	0.983	0.995	0.983	0.964
			0.90	0.884	0.899	0.882	0.763
		2.0	0.95	0.937	0.945	0.941	0.840
			0.99	0.983	0.991	0.980	0.935
			0.90	0.913	0.896	0.909	0.868
		0.5	0.95	0.956	0.948	0.952	0.933
			0.99	0.990	0.990	0.987	0.977
			0.90	0.899	0.907	0.895	0.851
	3	1.0	0.95	0.959	0.955	0.953	0.905
			0.99	0.989	0.993	0.988	0.971
			0.90	0.895	0.899	0.901	0.806
		2.0	0.95	0.947	0.950	0.946	0.869
			0.99	0.987	0.985	0.988	0.942
100			0.90	0.888	0.891	0.880	0.878
		0.5	0.95	0.946	0.945	0.938	0.931
			0.99	0.988	0.997	0.991	0.978
			0.90	0.891	0.904	0.898	0.849
	5	1.0	0.95	0.944	0.950	0.944	0.907
			0.99	0.985	0.991	0.988	0.967
			0.90	0.897	0.903	0.894	0.777
		2.0	0.95	0.946	0.948	0.947	0.846
			0.99	0.990	0.991	0.987	0.940

3. A Monte Carlo Study

3.1. Design of experiments

We consider a GLMM for clustered count responses defined by

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_i, \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, T_i, \tag{3.1}$$

where we let $\beta_0 = 1.5$, $\beta_1 = 1.0$, $\beta_2 = -1.0$. For each subject i the covariate variable x_{1ij} is randomly generated T_i times from uniform distribution over $(0, 1)$, and the covariate x_{2ij} is assigned to be 0 or 1 with respective probabilities $1/2$ for all elementary units of each subject. The x_{1ij} is sometimes called the within-subject variable and x_{2ij} is the subject-level covariate. The number of elementary

Table 3.2. Coverage probabilities of misspecified Poisson-normal GLMM when random effects distribution is $u_i \sim \mathsf{Gam}(\sigma^2, 1)$

n	T	σ	$1 - \alpha$	β_0	β_1	β_2	σ
50	3	0.5	0.90	0.920	0.916	0.876	0.469
			0.95	0.959	0.955	0.944	0.542
			0.99	0.993	0.987	0.989	0.674
		1.0	0.90	0.909	0.906	0.881	0.647
			0.95	0.940	0.953	0.939	0.712
			0.99	0.980	0.988	0.988	0.833
		2.0	0.90	0.890	0.901	0.907	0.655
			0.95	0.945	0.944	0.943	0.754
			0.99	0.981	0.992	0.983	0.866
	5	0.5	0.90	0.904	0.890	0.887	0.452
			0.95	0.958	0.946	0.953	0.503
			0.99	0.991	0.991	0.992	0.626
		1.0	0.90	0.890	0.899	0.875	0.619
			0.95	0.940	0.952	0.927	0.691
			0.99	0.982	0.992	0.980	0.804
		2.0	0.90	0.892	0.901	0.893	0.659
			0.95	0.943	0.953	0.940	0.730
			0.99	0.977	0.994	0.983	0.850
100	3	0.5	0.90	0.915	0.894	0.892	0.426
			0.95	0.963	0.947	0.943	0.499
			0.99	0.993	0.993	0.988	0.617
		1.0	0.90	0.909	0.909	0.883	0.576
			0.95	0.945	0.956	0.941	0.652
			0.99	0.993	0.993	0.987	0.783
		2.0	0.90	0.889	0.904	0.905	0.617
			0.95	0.945	0.955	0.941	0.720
			0.99	0.987	0.994	0.981	0.857
	5	0.5	0.90	0.897	0.895	0.882	0.399
			0.95	0.951	0.940	0.943	0.481
			0.99	0.988	0.987	0.991	0.600
		1.0	0.90	0.904	0.895	0.891	0.593
			0.95	0.947	0.945	0.946	0.663
			0.99	0.987	0.993	0.988	0.790
		2.0	0.90	0.896	0.909	0.900	0.620
			0.95	0.950	0.940	0.946	0.716
			0.99	0.989	0.990	0.984	0.862

units T_i is taken to be the same for all subjects, that is, $T_i = T$ for all $i = 1, 2, \ldots, n$. We consider two cases of $n = 50, 100$, and also we take $T = 3, 5$. As random effects distributions for u_i we consider the normal $N(0, \sigma^2)$, gamma $\mathsf{Gam}(\sigma^2, 1)$, normal mixtures $(1/2)N(-2, \sigma^2) + (1/2)N(2, \sigma^2)$, and exponential $\mathsf{Exp}(\sigma)$, where $\sigma = 0.5, 1.0, 2.0$. We standardize the generated random variates of u_i for all cases of random effects distributions.

By fitting the GLMM in (3.1) we obtain the MLEs of $\beta's$ and σ with their standard errors. The approximate confidence intervals for the parameters can also be obtained using the asymptotic normality of MLEs. We find the empirical coverage probability of confidence interval for each parameter by counting the number of confidence intervals containing the true parameter among 1,000 replications of the simulation study.

Table 3.3. Coverage probabilities of misspecified Poisson-normal GLMM when random effects distribution is $u_i \sim \text{Exp}(\sigma)$

n	T	σ	$1-\alpha$	β_0	β_1	β_2	σ
			0.90	0.929	0.912	0.919	0.637
		0.5	0.95	0.963	0.965	0.951	0.739
			0.99	0.981	0.990	0.990	0.878
			0.90	0.935	0.900	0.849	0.647
	3	1.0	0.95	0.945	0.945	0.935	0.732
			0.99	0.970	0.995	0.980	0.827
			0.90	0.883	0.895	0.856	0.513
		2.0	0.95	0.929	0.953	0.924	0.598
50			0.99	0.981	0.990	0.983	0.759
			0.90	0.890	0.872	0.912	0.716
		0.5	0.95	0.932	0.921	0.951	0.774
			0.99	0.979	0.983	0.985	0.861
			0.90	0.856	0.926	0.853	0.577
	5	1.0	0.95	0.935	0.983	0.881	0.634
			0.99	0.975	0.996	0.958	0.729
			0.90	0.800	0.829	0.854	0.587
		2.0	0.95	0.912	0.888	0.915	0.676
			0.99	0.972	0.973	0.973	0.700
			0.90	0.894	0.910	0.909	0.624
		0.5	0.95	0.894	0.910	0.909	0.624
			0.99	0.987	0.996	0.988	0.824
			0.90	0.943	0.878	0.902	0.567
	3	1.0	0.95	0.965	0.907	0.934	0.685
			0.99	0.995	0.964	0.994	0.834
			0.90	0.864	0.907	0.947	0.531
		2.0	0.95	0.953	0.934	0.971	0.602
100			0.99	0.981	0.997	0.997	0.736
			0.90	0.940	0.898	0.833	0.688
		0.5	0.95	0.956	0.949	0.872	0.728
			0.99	0.997	0.993	0.993	0.829
			0.90	0.897	0.901	0.883	0.593
	5	1.0	0.95	0.945	0.937	0.939	0.699
			0.99	0.979	0.988	0.985	0.832
			0.90	0.886	0.907	0.909	0.492
		2.0	0.95	0.937	0.963	0.944	0.574
			0.99	0.991	0.992	0.985	0.709

We define the empirical relative bias of true parameter θ_0 as

$$\frac{\bar{\hat{\theta}} - \theta_0}{\theta_0} \times 100, \tag{3.2}$$

where $\bar{\hat{\theta}}$ denotes the average of MLE $\hat{\theta}$ over 1,000 replications of simulations. The empirical relative bias approximates true relative bias for the parameter θ_0. Next, we investigate the impact on MLEs when response distribution is misspecified as Poisson whereas the true distribution is negative binomial having overdispersion.

3.2. Results

Table 3.1 shows that coverage probabilities for fixed parameters are moderately well attained when

Table 3.4. Coverage probabilities of misspecifed Poisson-normal GLMM when random effects distribution is $u_i \sim (1/2)N(-2, \sigma^2) + (1/2)N(2, \sigma^2)$

n	T	σ	$1-\alpha$	β_0	β_1	β_2	σ
50	3	0.5	0.90	0.873	0.878	0.889	0.940
			0.95	0.938	0.945	0.943	0.968
			0.99	0.983	0.988	0.986	0.993
		1.0	0.90	0.890	0.915	0.871	0.952
			0.95	0.945	0.958	0.933	0.978
			0.99	0.988	0.993	0.976	0.995
		2.0	0.90	0.883	0.897	0.893	0.832
			0.95	0.939	0.944	0.952	0.913
			0.99	0.988	0.993	0.984	0.970
	5	0.5	0.90	0.876	0.901	0.892	0.976
			0.95	0.943	0.949	0.938	0.990
			0.99	0.994	0.993	0.989	0.994
		1.0	0.90	0.870	0.901	0.892	0.954
			0.95	0.929	0.949	0.935	0.977
			0.99	0.985	1.000	0.990	0.997
		2.0	0.90	0.925	0.964	0.855	0.785
			0.95	0.926	0.965	0.893	0.856
			0.99	0.998	1.000	0.963	0.928
100	3	0.5	0.90	0.873	0.890	0.889	0.955
			0.95	0.923	0.951	0.932	0.982
			0.99	0.994	0.989	0.993	0.997
		1.0	0.90	0.913	0.858	0.857	0.928
			0.95	0.992	0.997	0.927	0.995
			0.99	1.000	0.999	0.999	1.000
		2.0	0.90	0.893	0.909	0.896	0.853
			0.95	0.946	0.956	0.942	0.922
			0.99	0.993	0.993	0.980	0.960
	5	0.5	0.90	0.856	0.880	0.880	0.988
			0.95	0.923	0.928	0.937	1.000
			0.99	0.985	0.979	0.994	1.000
		1.0	0.90	0.902	0.914	0.894	0.951
			0.95	0.949	0.955	0.942	0.991
			0.99	0.982	0.989	0.985	0.997
		2.0	0.90	0.870	0.912	0.877	0.826
			0.95	0.918	0.944	0.931	0.891
			0.99	0.991	0.988	0.969	0.978

the random effects distribution is correctly specified as normal; however, the coverage probabilities for the random component σ are slightly smaller than the nominal level. The performance of confidence intervals improves as sample size increases. Even when the true random effects distribution is $\mathrm{Gam}(\sigma^2, 1)$ of Table 3.2 and we incorrectly fit Poisson-normal GLMM, the nominal confidence levels for fixed parameters are well attained; however, the coverage errors of confidence intervals for σ are large irrespective of sample sizes. In the case of $\mathrm{Exp}(\sigma)$ random effects, we see similar results as given in Table 3.3. But as shown in Table 3.4 for the mixture of two normal distributions, the coverage probabilities for fixed effects and variance component have similar patterns as the normal random effects distribution of Table 3.1.

We find that the MLEs for fixed parameters are insensitive to misspecified random effects distribu-

Table 3.5. Relative biases of parameter estimates in Poisson-normal GLMM according to random effects distributions

u_i	n	T	σ	β_0	β_1	β_2	σ
$N(0, \sigma^2)$	50	3	0.5	−0.38	−0.27	−0.70	−3.73
			1.0	−0.74	0.03	−1.61	−3.08
			2.0	−1.36	0.04	−3.17	−3.06
		5	0.5	−0.31	−0.05	−0.54	−3.39
			1.0	−0.91	0.43	−1.46	−2.68
			2.0	−1.58	−0.12	−3.70	−2.80
	100	3	0.5	−0.11	0.31	0.22	−1.72
			1.0	0.05	−0.16	0.23	−1.33
			2.0	−0.08	0.06	0.57	−1.42
		5	0.5	0.02	0.02	0.20	−1.57
			1.0	−0.14	0.19	0.02	−1.24
			2.0	−0.07	−0.05	0.46	−1.21
$Gam(\sigma^2, 1)$	50	3	0.5	−1.33	0.24	0.47	7.24
			1.0	−1.51	0.02	3.29	4.01
			2.0	−4.68	−0.20	1.20	2.59
		5	0.5	−1.02	0.49	0.09	3.60
			1.0	−0.79	−0.06	2.59	1.47
			2.0	−3.46	0.03	−0.16	0.95
	100	3	0.5	−1.38	0.24	0.86	11.70
			1.0	−2.16	−0.33	2.36	5.68
			2.0	−4.69	0.02	2.38	4.00
		5	0.5	−0.83	0.20	0.74	7.19
			1.0	−1.66	0.09	1.29	3.04
			2.0	−3.25	−0.06	2.30	2.45
$Exp(\sigma)$	50	3	0.5	−1.17	0.67	2.06	8.33
			1.0	−0.75	−0.04	3.92	5.37
			2.0	−6.66	0.22	7.36	1.84
		5	0.5	0.58	−2.16	2.94	5.12
			1.0	−1.23	−0.43	1.17	4.14
			2.0	−4.92	−1.30	6.03	0.16
	100	3	0.5	−0.94	0.24	1.74	9.00
			1.0	−1.94	−1.33	0.59	6.87
			2.0	−7.23	−0.14	5.48	5.67
		5	0.5	−0.31	0.18	1.73	6.75
			1.0	−1.08	−0.06	2.24	3.07
			2.0	410.77	0.01	42.23	3336.73
$\frac{1}{2}N(-2, \sigma^2) + \frac{1}{2}N(2, \sigma^2)$	50	3	0.5	−0.61	0.80	−1.17	−3.35
			1.0	−0.84	0.80	−1.54	−0.20
			2.0	−0.42	−0.41	−3.61	−0.73
		5	0.5	0.28	0.10	0.65	−2.51
			1.0	−0.10	0.21	−0.45	−0.94
			2.0	−3.34	−0.99	−16.41	−1.31
	100	3	0.5	−0.11	0.06	−1.22	−1.78
			1.0	5.57	0.88	3.00	−0.08
			2.0	−0.72	−0.34	1.35	0.50
		5	0.5	−0.15	0.65	0.37	−1.32
			1.0	−1.33	0.15	0.79	1.26
			2.0	−1.52	0.01	1.01	−0.09

Table 3.6. Coverage probabilities of misspecified Poisson-normal GLMM when responses are negative-binomial with overdispersion parameter ϕ and $u_i \sim N(0, \sigma^2)$

n	ϕ	σ	$1 - \alpha$	β_0	β_1	β_2	σ
50	3	0.5	0.90	0.716	0.570	0.862	0.381
			0.95	0.823	0.667	0.918	0.550
			0.99	0.918	0.762	0.975	0.832
		1.0	0.90	0.742	0.380	0.904	0.825
			0.95	0.808	0.456	0.947	0.891
			0.99	0.942	0.566	0.984	0.971
		2.0	0.90	0.780	0.145	0.834	0.761
			0.95	0.853	0.146	0.951	0.878
			0.99	0.975	0.173	0.978	0.994
	5	0.5	0.90	0.543	0.254	0.878	0.001
			0.95	0.630	0.277	0.924	0.001
			0.99	0.761	0.395	0.966	0.034
		1.0	0.90	0.553	0.208	0.932	0.289
			0.95	0.670	0.269	0.965	0.391
			0.99	0.779	0.365	0.991	0.648
		2.0	0.90	0.646	0.078	0.894	0.756
			0.95	0.700	0.096	0.952	0.860
			0.99	0.816	0.127	0.988	0.967
100	3	0.5	0.90	0.705	0.535	0.899	0.165
			0.95	0.786	0.600	0.956	0.248
			0.99	0.898	0.731	0.997	0.488
		1.0	0.90	0.689	0.358	0.908	0.775
			0.95	0.752	0.391	0.961	0.881
			0.99	0.906	0.530	0.989	0.961
		2.0	0.90	0.687	0.134	0.900	0.774
			0.95	0.783	0.148	0.957	0.851
			0.99	0.908	0.188	0.985	0.938
	5	0.5	0.90	0.341	0.287	0.888	0.000
			0.95	0.409	0.333	0.929	0.000
			0.99	0.577	0.441	0.988	0.000
		1.0	0.90	0.426	0.128	0.901	0.061
			0.95	0.496	0.178	0.945	0.071
			0.99	0.643	0.244	0.981	0.160
		2.0	0.90	0.525	0.084	0.949	0.641
			0.95	0.602	0.091	0.980	0.757
			0.99	0.730	0.111	0.995	0.947

tions in general but the coverage probabilities are very sensitive to the misspecification of random effects of gamma and exponential distributions that are not symmetric but skewed. Biases of MLEs given in Table 3.5 have similar patterns as the coverage probabilities for each random effects distribution. Biases of fixed effects estimates are small in general when random effects distribution is correctly specified as normal or misspecified as a mixture of normal distributions. Biases for σ are a little more severe than those of fixed effects but decreases as sample size increases. When true random effects distribution is $\mathrm{Gam}(\sigma^2, 1)$ or $\mathrm{Exp}(\sigma)$ the performance of MLEs are not good. The random effects u_i's generated from $\mathrm{Exp}(\sigma)$ tend to have large variability as the σ increases; subsequently, the asymptotic properties of MLEs seem to worsen in respect to low coverage probabilities and the relative biases given in Table 3.3 and Table 3.5, respectively. We also comment that there

Table 3.7. Relative biases of misspecified Poisson-normal GLMM when responses are negative-binomial with overdispersion parameter ϕ and $u_i \sim N(0, \sigma^2)$

n	ϕ	σ	β_0	β_1	β_2	σ
		0.5	-6.69	1.59	-0.80	28.99
	0.5	1.0	-4.86	-2.68	2.00	5.03
		2.0	0.07	-14.71	-8.67	-1.92
		0.5	-10.60	-3.81	-5.28	50.59
50	1.0	1.0	-12.19	-7.20	2.68	9.68
		2.0	-41.61	56.76	-0.99	7.31
		0.5	-20.76	-7.88	2.25	96.93
	2.0	1.0	-28.10	1.18	-3.24	31.60
		2.0	-28.92	3.71	2.09	7.87
		0.5	-5.55	0.36	1.44	28.73
	0.5	1.0	-6.46	-0.12	-1.61	6.70
		2.0	-7.06	-0.62	-1.95	0.66
		0.5	-12.09	-0.82	-2.49	56.24
100	1.0	1.0	-12.94	0.78	-7.61	15.94
		2.0	-14.26	-0.54	-0.47	2.90
		0.5	-24.97	1.58	-1.79	101.25
	2.0	1.0	-25.91	0.94	-1.47	33.83
		2.0	-26.82	-4.74	3.27	8.24

has been very few cases of no convergence among 1,000 replications of simulations fitting GLMM when we misspecify as $N(0, \sigma^2)$ whereas the true distribution is $Gam(\sigma^2, 1)$ or $Exp(\sigma)$.

We should be cautious that the highly biased estimate of variance component makes it difficult to evaluate the properties of the fixed parameters as well. The estimation of β_2 and σ is subject to between-individual variation; therefore, we expect that the misspecification of the random effects distribution affects the quality of these estimates.

However, as discussed by Heagerty and Kurland (2001), and Litiere, *et al.* (2007), the parameters of within-subjects covariates would be less affected by the misspecification of random effects distribution. As a second experiment we simulated negative binomial responses with overdispersion parameter ϕ. Table 3.6 shows that coverage probabilities are not attained at all even for fixed parameters except β_2, the coefficient of subject-level covariate x_{2ij}. They are extremely low compared to nominal levels when $\sigma = 0.5$, and they improve as σ increases. We also note that as ϕ increases the performance of confidence interval deteriorates. Table 3.7 shows the biases of estimates when Poisson-normal GLMM has been fitted for the overdispersed counts data.

4. Summary and Further Research

The generalized linear mixed model is very useful for fitting categorical responses such as binomial or Poisson counts of clustered data. In GLMM the random effects distribution is assumed to be normal, and most commercial software routinely fit GLMM only for the normal random effects. We investigated the sensitivity or robustness of the MLEs of GLMM for the count responses when the normality assumption is violated. Even when the assumed model is misspecified the MLE is consistent to the unknown parameter θ_* which is determined from the Kullback-Leibler information criteria discussed in Section 2.2. Our interest has been focused to investigate the properties of MLEs under misspecifications. Through a Monte Carlo study we discussed performances of MLEs

in terms of coverage probability and bias when pseudo random variates were respectively generated from the distribution of normal, gamma, exponential, and mixture of two normal distributions. Furthermore, we considered the effects on the MLEs when the response distribution of counts data is misspecified as Poisson when the responses were generated from the negative binomial distribution with overdispersion.

The simulation result shows that the performance of MLEs for fixed parameters are moderately good even when the random effects distribution is incorrectly specified; however, the results for the variance component σ are very poor when u_i are generated from gamma or exponential distribution that is not symmetric but skewed. This fact denotes that the estimates of fixed parameters are relatively robust to the misspecification of random effects distribution but the estimate of variance component is very sensitive to asymmetric random effects distributions. The results of the second experiment of misspecified response distribution show also that the performance of MLEs are poor, in particular, when σ is small. The performance deteriorates as overdispersion increases.

It would be interesting to investigate the performance of MLEs of fixed parameters corresponding to between-subjects covariates or within-subjects covariates. We further need to study the performance of MLEs in a very small number of subjects with a relatively large number of elementary units for each subject.

References

Agresti, A. (2002). *Categorical Data Analysis*, Second Ed., Wiley, New York.

Alonso, A., Litiere, S. and Molenberghs, G. (2010). Testing for misspecification in generalized linear mixed models, *Biostatistics*, **11**, 771–786.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second Ed., London, Chapman and Halls.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalized linear mixed models, *Biometrika*, **88**, 9–25.

Litiere, S., Alonso, A. and Molenberghs, G. (2007). Type I and Type II error under random-effects misspecification in generalized linear mixed models, *Biometrics*, **63**, 1038–1044.

Litiere, S., Alonso, A. and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models, *Statistics in Medicine*, **27**, 3125–3144.

Molenberghs, G., Declerck, L. and Aerts, M. (1998). Misspecifying the likelihood for clustered binary data, *Computational Statistics and Data Analysis*, **26**, 327–349.

Neuhaus, J. M., Hauck, W. W. and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models, *Biometrika*, **79**, 755–762.

Whites, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–26.