

Impact of Sample Sizes on the Accuracy of Estimates for a Two-level Logistic Regression Model

Roll No: FH-033-015
Session: 2018-19

A project report submitted as a partial requirement for the
degree of B.S. (Honors) in Applied Statistics



Institute of Statistical Research and Training
University of Dhaka
Bangladesh
August 2019

Approved for the Division
(Applied Statistics)

Noor Jahan Akter

Abstract

In a multilevel framework an important problem is calculating an adequate sample size that generates accurate estimates. Several researches have investigated the behavior of estimates in finite samples, particularly for continuous dependent variables. On the other hand, binary response multilevel models have been investigated less widely. In this research we investigate the sample behavior of estimates in a binary response multilevel model. Particularly, we analyze the influence of different factors on the accuracy of estimates and their profile likelihood confidence intervals for a 2-level logistic regression model, through a Monte Carlo simulation study. We investigate the hypothesis of: (a) different level-1 sample sizes; (b) different level-2 sample sizes; (c) different intra-class correlation coefficients. We investigate the bias of estimates by relative bias and, through a non-coverage indicator, the accuracy of the confidence interval. In all instances we have examined, the point estimates are unbiased (even with very small sample sizes), while the variance components are always underestimated. Despite some exceptions for the variance components, the profile likelihood confidence interval performs very well in all simulated conditions.

Table of Contents

| | |
|--|-----------|
| Preface | 1 |
| Chapter 1: Introduction | 2 |
| 1.1 Background | 2 |
| 1.2 Literature review | 3 |
| 1.3 Motivation of the study | 5 |
| 1.4 Objectives of the study | 5 |
| 1.5 Organization of the study | 6 |
| Chapter 2: Methodology | 7 |
| 2.1 Multilevel model | 7 |
| 2.2 The multilevel linear regression model | 7 |
| 2.3 The multilevel logistic regression model | 8 |
| 2.4 Parameter estimation in multilevel logistic regression model | 10 |
| 2.5 Penalized quasi-likelihood | 12 |
| 2.6 Adaptive Gaussian quadrature | 12 |
| 2.7 Laplace approximation | 13 |
| 2.8 Confidence interval of the estimates | 14 |
| 2.8.1 Wald-type confidence interval | 14 |
| 2.8.2 Profile likelihood confidence interval | 14 |
| Chapter 3: Simulation and analysis techniques | 16 |
| 3.1 Simulation model and procedure | 16 |
| 3.2 Analysis techniques | 18 |
| 3.2.1 Percentage relative bias | 18 |
| 3.2.2 Non-coverage rate of confidence interval | 18 |

| | |
|---|-----------|
| Chapter 4: The simulation results | 19 |
| 4.1 Rate of model convergence | 19 |
| 4.2 The bias in point estimates | 19 |
| 4.3 Non-coverage of confidence interval | 21 |
| Chapter 5: Conclusion | 25 |
| Appendix A: R codes used in this study | 27 |
| References | 34 |

List of Tables

| | | |
|-----|---|----|
| 4.1 | The effect of number of groups, group size, and ICC on the percentage relative bias ($\frac{\hat{\theta}-\theta}{\theta} \times 100$) of the estimates. | 20 |
| 4.2 | Non-coverage of the 95% confidence interval by number of groups. . . | 22 |
| 4.3 | Non-coverage of the 95% confidence interval by group size. | 22 |
| 4.4 | Non-coverage of the 95% confidence interval by Intra-class correlation (ICC). | 22 |
| 4.5 | The effect of number of groups, group size, and ICC on the non-coverage of the 95% profile confidence interval. | 23 |

List of Figures

| | | |
|-----|---|----|
| 4.1 | Dot plots of non-coverage rates of the profile likelihood confidence interval over different parameter estimates. | 24 |
|-----|---|----|

Acknowledgements

First of all, I am grateful to the one true God for the good health and wellbeing that were necessary to complete this project. I wish to express my sincere gratitude to my project supervisor Ms. Noor Jahan Akter. I would also like to thank my friends and classmates for offering me advice, and supporting me through this entire process. Special thanks to my friends and brothers from Fazlul Huq Muslim Hall. Without their help and support, I could not have made it this far. Finally, I must express my very profound gratitude to my parents for providing me with unfailing support throughout my years of study. Thank you.

Declaration

I, Md. Rasel Biswas hereby formally declare that I have written the submitted Master's thesis entirely by myself without anyone else's assistance. Where I have drawn on literature or other sources, either in direct quotes, or in paraphrasing such material, I have referenced the original author or authors and the source in which it appeared.

Md. Rasel Biswas
Dhaka, August 2019

Dedication

To my parents.

Preface

This is Rasel Biswas's final year project dissertation for the B.S. (honors) degree at the University of Dhaka. The gitbook version can be read via the following link:
<https://rasel.rbind.io/project450>

Chapter 1

Introduction

1.1 Background

Most of the statistical inference is based on replicated observations of units of analysis of one type (e.g., a sample of individuals, countries, or schools). The analysis of such observations usually is based on the assumption that either the sampled units themselves or the corresponding residuals in some statistical model are independent and identically distributed. However, many kinds of data, including social surveys, have a hierarchical or clustered structure, which includes units of analysis of more than one type. For instance, in educational research, researchers often study how students' academic outcomes are associated with the characteristics of their families, classrooms, and schools. The student test scores are frequently modeled as the result of a combination of individual-level characteristics such as student socioeconomic status (SES) and IQ score, classroom level characteristics, such as the number of students in a classroom, and school-level characteristics, such as the expertise of the teachers or location of the school (Urban, Semi-urban or Rural). In this study, pupils, teachers, classrooms, and schools might all be important units of analysis. Frequently, but by no means always, units of analysis of different types are hierarchically nested (e.g., pupils are nested in classrooms, which, in turn, are nested in schools). Other examples involve organizational research, with employees nested within firms; cross-national comparative research with individuals nested within countries; medical research with patients nested within clinics; and longitudinal research with repeated measures nested within individuals. Since these analyses include the units of analysis of more than one type, the traditional statistical techniques (i.e., ordinary least squares regression) can

not be used in these cases. Multilevel analysis is a general term referring to statistical methods appropriate for the analysis of data sets comprising several types of units of analysis. The levels in the multilevel analysis are another name for the different types of unit of analysis (Goldstein, 2003).

Multilevel analysis, originally developed in the fields of education, sociology, and demography, has received increasing attention in other fields like public health and epidemiology over the past few years. Multilevel analysis has become popular in many fields that have hierarchically organized data in units of analysis at two or more levels. Multilevel analysis can be used either to adjust for the dependency of the observations within clusters by using variables at higher levels, or assessing the impact of higher-level characteristics on the outcome variables after controlling for individual characteristics at a base level (Snijders & Bosker, 1999).

An important problem that occurs in the practice of the multilevel model is the choice of sample sizes at different levels. Applied practitioners designing a research study are often concerned with the required sample size necessary to estimate unbiased fixed effects and variance components, avoid model convergence issues, and yield accurate confidence intervals around point estimates. Insufficient sample sizes may lead to biased estimates and/or significance tests with low statistical power (Hox, 2010). Since the 1990s, sample size considerations in multilevel studies have become the topic of a huge amount of contributions. There are plenty of researches on sample size determination for continuous response multilevel model. However, few studies have been carried out to date that has focused on determining sample sizes for the binary response multilevel model.

1.2 Literature review

There is a somewhat wide literature about simulation and studies on this issue for continuous response multilevel models. The robustness issue and the choice of sample size in the multilevel regression model for continuous outcome variable have been studied by several authors (Akter & Khan, 2018; Clarke & Wheaton, 2007; Hox, 1998; Kreft, 1996; Maas & Hox, 2004, 2005; Mok, 1995). The overall conclusion from these studies is that the estimates of the regression coefficients are nearly unbiased, however, the variance components tend to be biased downward (underestimated) when the number of level 2 units is small (e.g., less than 30) (Maas & Hox, 2004, 2005). These

simulation studies also indicate that a larger number of groups is more important than a larger number of individuals per group (Hox, 2010; Maas & Hox, 2004). For two-stage hierarchical linear models, Akter & Khan (2018) recommend a minimum of 30 groups with 15 units per each group for drawing a valid conclusion about the population.

However, few studies have assessed the accuracy of estimates, sample size or power analysis in multilevel regression for a binary outcome variable. Austin (2005) used Monte Carlo simulation methods to examine the impact of misspecification of the distribution of the random effects on estimation of and inference about both the fixed effects and the random effects in multilevel logistic regression models. He concluded that estimation and inference concerning the fixed effects were insensitive to misspecification of the distribution of the random effects, however, estimation and inferences concerning the random effects were affected by model misspecification. Sastry, Ghosh-Dastidar, Adams, & Pebley (2006) calculate power and sample size in multilevel logistic regression models for their survey of children, families, and communities in Los Angeles. Based on simulation studies, they decided to sample 65 groups each of size 50.

Moineddin, Matheson, & Glazier (2007) perform some simulation studies on a random slope logistic model using the AGQ with automatic selection of the number of quadrature points as the estimation procedure. Differently from the continuous case, they suggest a “50/50” rule (a minimum of 50 groups with at least 50 units per group) for achieving sufficient accuracy. The estimates of all random components, as well as their standard errors, are underestimated, even with very large sample size. They also investigate how the prevalence of the outcome affects the accuracy of individual-level parameter estimates, finding that group size should be properly adjusted in case of low prevalent outcomes. Furthermore, their results signal some rather low percentages of model convergence in some studies, particularly with low group sizes.

Paccagnella (2011) found that the regression parameter estimates appear unbiased even with small sample size, while the variance components are underestimated. The bias decreases as the number of groups increases. Like continuous response multilevel models, the accuracy of the standard errors of the regression parameter estimates is achieved with a number of groups equal to 50. Unlike continuous response models, the accuracy of the standard errors of variance estimates needs a very large number of groups. The simulation studies of Moineddin et al. (2007) and Paccagnella (2011)

investigated the effect of sample size based on the asymptotic theory, i.e., they assumed that the sample size is sufficiently large to conduct the significance test. Even though the sample size is large enough, testing a variance component using their standard error is woefully inadequate (Bates, 2010).

1.3 Motivation of the study

A Wald-type confidence interval construction is based on the asymptotic normality of the maximum likelihood estimate. It is well known, however, that properties of estimates in small samples can be very different from the asymptotic properties. A more robust construction of confidence interval is profile likelihood-based confidence interval. Profile likelihood works well in both situations when the sample size is small and variance components are to be estimated (Venzon & Moolgavkar, 1988). However, in most of the studies reviewed in Section 1.2, to recommend a minimum sample size for accurate confidence intervals, they examined the accuracy of the Wald-type confidence interval.

In addition to this, a recent simulation study of Handayani, Notodiputro, Sadik, & Kurnia (2017) show that for obtaining maximum likelihood estimates in generalized linear mixed models, the Laplace approximation method produces better estimates when compared to penalized quasi-likelihood (PQL) and adaptive Gaussian quadrature (AGQ) approximation methods in terms of their relative biases and mean square errors.

We are unaware of any simulation studies to date that have investigated the impact of sample sizes on the accuracy of parameter estimates and their profile likelihood confidence intervals, using Laplace approximation estimation procedure, for a two-level logistic regression model.

1.4 Objectives of the study

This study focuses on the impact of sample sizes on the accuracy of parameter estimates (fixed effects and variance components) and their profile likelihood confidence intervals. The accuracy of the parameter estimates is assessed by the percentage relative bias. The accuracy of the confidence interval, for each parameter, is examined by analyzing the observed coverage of the 95% confidence interval. Based on the findings, this simulation study attempts to recommend the minimum sample size necessary to

estimate unbiased fixed effects and variance components, avoid model convergence issues, and yield accurate confidence intervals around point estimates for a two-level logistic regression model.

1.5 Organization of the study

The organization of this project is as follows. Chapter 1 contains the background, reviews of the literature, and objectives. Chapter 2 describes multilevel logistic regression model, parameter estimation techniques, and steps to construct profile likelihood confidence interval. In Chapter 3, Simulation model and simulation process are described, as well as the analysis techniques are discussed. Chapter 4 provides the simulation results. Finally, Chapter 5 summarizes the project work and conclude our objectives and results.

Chapter 2

Methodology

2.1 Multilevel model

Multilevel models (also known as hierarchical linear models, linear mixed-effects models, mixed models, nested data models, random coefficient, random-effects models, random parameter models, or split-plot designs) are statistical models of parameters that vary at more than one level (Raudenbush & Bryk, 2002). Multilevel models are appropriate particularly for research designs where data for the individuals are organized at more than one level (i.e., nested data or grouped data). Multilevel models can be used on data with many levels, although two-level models are the most common. In a two-level model, the units of analysis are usually individuals (at a lower level) who are nested within contextual/aggregate units, sometimes called cluster or group (at a higher level). This project deals only with two-level models.

2.2 The multilevel linear regression model

Assume that we have data from J groups, with different numbers of individuals, n_j , in each group. Let, Y_{ij} denote a continuous outcome variable on the i^{th} individual in the j^{th} group. Also, assume that we have one explanatory variable X_{ij} on the individual level (level-1), and one explanatory variable Z_j on the group-level (level-2).

In order to model the data, we consider a two-level linear regression model with one explanatory variable at each level. The model is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \quad (2.1)$$

where β_{0j} is the group specific intercept, β_{1j} is the group specific effect of the individual level variable X_{ij} , and e_{ij} are the individual level residuals. The variation of the group specific regression coefficients β_{0j} and β_{1j} are modeled as a function of group level (or higher level) variables as follows:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j},\end{aligned}\tag{2.2}$$

where γ_{00} is the common intercept across groups, γ_{01} is the effect of the group level predictor on the group-specific intercepts, γ_{10} is the common slope associated with the individual level variable across groups, and γ_{11} is the group level predictor on the group-specific slopes. The individual-level residuals e_{ij} are assumed to have a normal distribution with mean zero and variance σ_e^2 . The group-level residuals u_{0j} and u_{1j} are assumed to have a multivariate normal distribution with expectation zero, and to be independent from the residual errors e_{ij} . the variances of the residual errors u_{0j} and u_{1j} are specified as σ_{u0}^2 and σ_{u1}^2 . That is,

$$e_{ij} \sim N(0, \sigma_e^2),$$

and

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{u1}^2 \end{bmatrix} \right).$$

This model can be written as one single regression model by substituting Equations (2.2) into the Equation (2.1). After substituting and rearranging terms the model becomes as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij} + e_{ij}.\tag{2.3}$$

2.3 The multilevel logistic regression model

Let Y_{ij} denote a binary outcome variable, coded 0 or 1, on the i^{th} individual in the j^{th} group. Also p_{ij} be the probability that the outcome variable equals 1, i.e., p_{ij} is the probability that the i^{th} individual in j^{th} group will experience the outcome. Here Y_{ij} follows a Bernoulli distribution. Then in order to model the data, p_{ij} is modeled using

the link function, ‘logit’. So for the binary outcome variable the model is as follows:

$$\text{logit}(p_{ij}) = \pi_{0j} + \pi_{1j}X_{ij}, \quad (2.4)$$

where π_{0j} is the group specific intercept, and π_{1j} is the group specific effect of the individual level variable X_{ij} . The variation of the group specific regression coefficients are modeled as a function of group level (or higher level) variables as follows:

$$\begin{aligned} \pi_{0j} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \\ \pi_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j}, \end{aligned} \quad (2.5)$$

where

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{u1}^2 \end{bmatrix} \right).$$

The Model (2.4) can be rewritten as the following single equation:

$$\text{logit}(p_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}. \quad (2.6)$$

Equation (2.6) consists of two parts. The segment $\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j$ contains all the fixed coefficients; it is called the fixed effect (or deterministic) part of the model. The segment $u_{0j} + u_{1j}X_{ij}$ is called the random (or stochastic) part of the model because it contains all the random error terms. Even if the analysis includes only the individual level variable, standard multivariate models are not appropriate for the grouped data since grouped data violate a crucial assumption of the independence of all observations. **Intra-class correlation (ICC)** is used as a mean to quantify the amount of dependence caused by grouping. ICC represents the proportion of the total observed individual variation in the outcome that can be explained by the grouping structure in the population. In other words, $\text{ICC} = 0$ indicates perfect independence of residuals: the observations do not depend on the group membership. On the contrary, $\text{ICC} = 1$ indicates perfect interdependence of residuals: the observations only vary between groups (Sommet & Morselli, 2017). When the ICC is not different from zero or negligible, one could consider running traditional one-level regression analysis rather than multilevel regression. The intraclass correlation can also be interpreted as the expected correlation between two randomly drawn units that are in the same group.

The Intra-class correlation (ICC) can be estimated by specifying a fully uncondi-

tional multilevel logistic model:

$$\text{logit}(p_{ij}) = \gamma_{00} + u_{0j}. \quad (2.7)$$

The ICC for the logistic regression model is defined as

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}, \quad (2.8)$$

where σ_{u0}^2 is the variance of the random intercept in the fully unconditional model and σ_e^2 is the variance of the residuals on the first level. However, for logistic regression, there is no direct estimation of the variance of the residuals σ_e^2 on the first level. Among different procedures, ‘latent variable approach’ has become the most widely used procedure for computing ICC in a multilevel logistic model. In this approach, the observed binary response is considered to represent a thresholded continuous variable where we observe 0, when the value of the latent variable is below the threshold, and 1 otherwise (e.g., pass/fail on a test is a binary representation of an underlying continuous latent variable test score where the threshold is the pass mark). In a ‘logit’ model, the underlying latent variable has a logistic distribution. We know that the logistic distribution has variance $\frac{\pi^2}{3}$. We can then take this as the level 1 variance so that now both the level 1 and 2 variances are on the same scale since the level 2 variance is measured on the logistic scale (Guo & Zhao, 2000).

2.4 Parameter estimation in multilevel logistic regression model

From previous description recall that, the number of groups in our data is J , and n_j is the number of observations within the j^{th} group. Now suppose that the number of regressors in our analysis is p (including the intercept) and there are r random effects where $p \geq r$. Then the multilevel linear regression model in Equation (2.3), for the j^{th} group, can be written as:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j, \quad (2.9)$$

where \mathbf{Y}_j is an $n_j \times 1$ vector of responses for group j , \mathbf{X}_j is an $n_j \times p$ design matrix for the regressors in group j , $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed effects coefficients, \mathbf{Z}_j is

an $n_j \times r$ design matrix for the random effects of group j , \mathbf{u}_j is a $r \times 1$ vector of unknown random effects for group j where $\mathbf{u}_j \sim MVN(\mathbf{0}, \mathbf{D})$, $\boldsymbol{\varepsilon}_j$ is a vector of unobserved errors of the observations in group j where $\boldsymbol{\varepsilon}_j \sim MVN(\mathbf{0}, \mathbf{R}_j)$ and $\text{Cov}(\mathbf{u}_j, \boldsymbol{\varepsilon}_j) = \mathbf{0}$.

In the multilevel logistic regression model, the probability of outcome variable Y_j , conditional on the group-specific effects u_j , can be written as

$$P(Y_j = 1|\mathbf{u}_j) = \frac{\exp(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j)}{1 + \exp(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j)}, \quad (2.10)$$

or as

$$P(Y_j = 0|\mathbf{u}_j) = \frac{1}{1 + \exp(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j)}. \quad (2.11)$$

We assume that responses in each group are independent after conditioning on the random effects (i.e., $\mathbf{R}_j = \sigma^2\mathbf{I}$). Then the conditional probability of \mathbf{Y}_j can be written as

$$f(\mathbf{Y}_j|\mathbf{u}_j; \boldsymbol{\beta}) = \prod_{i=1}^{n_j} P(Y_{ij} = 1|\mathbf{u}_j)^{Y_{ij}} P(Y_{ij} = 0|\mathbf{u}_j)^{1-Y_{ij}}, \quad (2.12)$$

where i is an index for individuals within clusters, i.e., Y_{ij} is the i^{th} individual within the j^{th} group. Because the random effects are unobserved, inferences for the fixed effects $\boldsymbol{\beta}$ and the covariance matrix of the random effects \mathbf{D} are estimated by integrating over the random effects, \mathbf{u}_j . The result is the ‘marginal likelihood function’ which is written as

$$L(\boldsymbol{\beta}, \mathbf{D}) = \prod_{j=1}^J \int f(\mathbf{y}_j|\mathbf{u}_j; \boldsymbol{\beta}) r(\mathbf{u}_j; \mathbf{D}) d\mathbf{u}_j, \quad (2.13)$$

where $r(\mathbf{u}_j; \mathbf{D})$ is a probability distribution of \mathbf{u}_j .

So, the expression for the marginal likelihood of a multilevel model is an integral over the random effects space. The calculation involves high dimensional integrals. For a linear multilevel model, this integral can be evaluated exactly. For a generalized linear multilevel model the integral must be approximated since unlike the linear multilevel model, the likelihood does not have a closed-form solution because of the inherent non-linearity of the model (McNeish, 2016). This leaves researchers with two broad approaches to estimate the model (1) linearly approximate the model so that the likelihood function does have a closed-form (e.g., penalized quasi-likelihood) or (2) retain the non-linearity of the model and approximate the likelihood function (e.g., Gaussian quadrature, Laplace approximation).

The three most common techniques for estimating multilevel logistic models are

briefly discussed here:

2.5 Penalized quasi-likelihood

Penalized quasi-likelihood (PQL) approximates the model by linearizing its non-linear components rather than approximating the integral of the likelihood function (Breslow & Clayton, 1993). This can be done with a double iterative algorithm. At first, the fixed effects (β) and random effects (\mathbf{u}_j) are estimated (either fitting a single-level logistic regression model or the estimates can be user specified). These initial estimated values are referred to as $\tilde{\beta}$ and $\tilde{\mathbf{u}}_j$. Then, conditional on the estimates of β and \mathbf{u}_j , the variance components (\mathbf{D} and \mathbf{R}_j) are estimated. Then, based on the estimates of \mathbf{D} and \mathbf{R}_j , β and \mathbf{u}_j are updated and the cycle continues until the differences between the estimates between two successive iterations are sufficiently small (Goldstein & Rasbash, 1996).

A basic advantage of PQL over other computational methods for multilevel logistic models is its computational efficiency. In terms of computational speed, PQL is often the most efficient relative to other estimation methods. Therefore, PQL estimation is sometimes advocated as a starting value for other procedures and for exploratory reasons. Another advantage is that for complex models (e.g., having a large number of random effects and/or multiple hierarchies) the model may still be estimated by PQL, while other estimation methods fail. However, one big disadvantage of this method is the parameter estimated from PQL are negatively biased (Breslow & Lin, 1995).

2.6 Adaptive Gaussian quadrature

Gaussian quadrature (GQ) is a numerical approximation method that partitions the marginal likelihood function from Equation @ref{eq:likelihood} into multiple components and then evaluates the integral by a weighted sum over the component partitions. As more quadrature points are selected, the approximation becomes more accurate. In traditional GQ, the quadrature points are centered around zero. Adaptive Gaussian quadrature (AGQ) centers the quadrature points about the mode of the marginal likelihood function which is especially advantageous when the mode is distant from zero (Lesaffre & Spiessens, 2001).

In AGQ method, the approximate ‘marginal likelihood’ is calculated by

$$L(\boldsymbol{\beta}, \mathbf{D}) \approx \prod_{j=1}^J \sum_{q=1}^Q f(\mathbf{Y}_j | \mathbf{u}_j = v_q) w_q, \quad (2.14)$$

where Q is the number of quadrature points (which are user selected), so the number of partitions is $Q + 1$, v_q is the q^{th} evaluation point and w_q is the related weight (Fitzmaurice, Laird, & Ware, 2011). One main disadvantage of AGQ is the computational burden since the number of computations increases exponentially as the number of random effects increases.

2.7 Laplace approximation

Another approach is the Laplace approximation. The goal of the Laplace approximation is to provide an approximation for the marginal likelihood function so that integration can be performed. The Laplace approximation uses Taylor series expansions to approximate the integrand rather than computing the integral with numerical methods as with AGQ so that the integral will have a closed-form solution.

The marginal likelihood function in Equation (2.13) can alternatively be written as

$$L(\boldsymbol{\beta}, \mathbf{D}) = (2\pi)^{-k/2} |\mathbf{D}|^{-1/2} \int \exp[h(\mathbf{u}_j)] d\mathbf{u}_j, \quad (2.15)$$

where

$$h(\mathbf{u}_j) = \log f(\mathbf{Y}_j | \mathbf{u}_j; \boldsymbol{\beta}) - \frac{1}{2} (\mathbf{u}_j^T \mathbf{D}^{-1} \mathbf{u}_j). \quad (2.16)$$

A second-order Taylor series expansion is applied to $\exp[h(\mathbf{u}_j)]$ about the mode of \mathbf{u}_j (denoted $\tilde{\mathbf{u}}_j$) such that

$$\exp[h(\mathbf{u}_j)] = \exp \left[h(\tilde{\mathbf{u}}_j) + \frac{1}{2} (\mathbf{u}_j - \tilde{\mathbf{u}}_j)^T \frac{\partial^2 h}{\partial \mathbf{u}_j \partial \mathbf{u}_j^T} (\mathbf{u}_j - \tilde{\mathbf{u}}_j) + K_j \right], \quad (2.17)$$

where K is the remainder which is ignored because it approaches zero as the group size increases. After expansion, the integrand has a closed-form and the integral can be evaluated (Raudenbush, Yang, & Yosef, 2000). Laplace method yields more accurate fixed and variance component estimates as sample sizes increase, particularly compared to PQL methods (Breslow & Lin, 1995). Using a single point in AGQ method is equivalent to the Laplace approximation. The number of points greater

than 1 produce greater accuracy in the evaluation of the likelihood. However, the larger the number of points, the more computationally intensive (and restrictive) is the estimation procedure.

2.8 Confidence interval of the estimates

2.8.1 Wald-type confidence interval

The $(1 - \alpha) \cdot 100\%$ Wald-type confidence interval is found as

$$(\hat{\theta} - z_{\alpha/2} \cdot SE(\hat{\theta}), \hat{\theta} + z_{\alpha/2} \cdot SE(\hat{\theta})), \quad (2.18)$$

where $\hat{\theta}$ is the estimate of the parameter θ , and $SE(\hat{\theta})$ is the standard error of the estimate $\hat{\theta}$. The $z_{\alpha/2}$ is the $(1 - \alpha/2) \cdot 100\%$ percentile of the standard normal distribution, which is the sampling distribution of the Wald statistic in repeated samples, when the sample size is large.

2.8.2 Profile likelihood confidence interval

The main idea of profile likelihood is to invert a likelihood-ratio test to obtain a CI for the parameter in question (Venzon & Moolgavkar, 1988). Consider a statistical model with parameters θ and δ where θ is the parameter of interest and δ is the (vector of) additional parameter(s) in the model. We denote the likelihood function by $L(\theta, \delta)$, and the maximum likelihood (ML) estimates by (θ^*, δ^*) . For the hypothesis $H_0 : \theta = \theta_0$ (where θ_0 is a fixed value), the likelihood ratio test statistic (G^2) equals the drop in $2 \ln L$ between the full model and the reduced model with θ fixed at θ_0 , i.e.,

$$G^2 = 2 (\ln L(\theta^*, \delta^*) - \ln L(\theta_0, \delta_0^*)), \quad (2.19)$$

where δ_0^* is the ML estimate of the reduced model.

Alternatively, we may express the test statistic G^2 in terms of the profile likelihood function L_1 for the parameter θ which is obtained from the usual likelihood function by maximizing over the parameter δ , i.e., $L_1(\theta) = \max_{\delta} L(\theta, \delta)$. Then we have

$$G^2 = 2 (\ln L_1(\theta^*) - \ln L_1(\theta_0)). \quad (2.20)$$

A 95% CI for θ consists of those values of θ_0 for which the test is non-significant at significance level 0.05; this is the case when G^2 does not exceed 3.84 (95%-percentile of the $\chi^2(1)$ distribution). Thus, the Confidence Interval consists of the θ_0 -values for which $G^2 = 2(\ln L_1(\theta^*) - \ln L_1(\theta_0)) \leq 3.84$, or, $\ln L_1(\theta_0) \geq \ln L(\theta^*, \delta^*) - 1.92$. For a confidence interval with coverage $(1 - \alpha) * 100\%$, use instead the $(1 - \alpha)$ -percentile of the $\chi^2(1)$ distribution.

Computation of a profile likelihood confidence interval follows some steps. For simplicity, we consider only the lower bound of the CI (the upper bound is similar) and assume the profile likelihood function to be an increasing function to the left of its maximum. As a start, compute the ML estimates (θ^*, δ^*) and the corresponding log-likelihood value. Then proceed by the following steps:

1. Compute a ‘reasonable’ lower bound θ' for the lower confidence limit (e.g., $\theta^* - 5 \text{ SE}(\theta^*)$, or 0.0001 if θ -values are restricted to be > 0).
2. Define a grid of values ranging from θ' to θ^* (e.g., 100 equidistant points).
3. For each grid value θ_i , compute the profile log-likelihood value $\ln L_1(\theta_i)$ by maximizing the $\ln L(\theta_i, \delta)$ over δ -values (a standard analysis allowing θ to be fixed at θ_i may apply).
4. Take as the lower bound (θ_L) of the 95% CI the smallest θ_i -value for which it holds that $\ln L_1(\theta_i) \geq \ln L(\theta^*, \delta^*) - 1.92$.
5. If necessary, refine or extend the grid of values around θ_L to obtain greater accuracy.

For repeated computations, one may replace the crude search over a grid of values by a systematic search procedure (e.g., bisection of the interval from θ' to θ^*).

Chapter 3

Simulation and analysis techniques

3.1 Simulation model and procedure

For this simulation study, we focus on the following two-level model, with one explanatory variable at level 1 (individual level) and one explanatory variable at level 2 (group level):

$$\text{logit}(p_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}. \quad (3.1)$$

The accuracy of the estimates can be affected by the size of the intra-class correlation coefficient (ICC) (Goldstein, 2003). Therefore, in our simulation, we have varied not only the sample size at the individual and group level but also the ICC.

Following the simulation conditions used by Maas & Hox (2005) and Moineddin et al. (2007), we set the following conditions for our simulation study:

1. the number of groups (J) is set at 30, 50, and 100,
2. the number of individuals per group (n_j) is set at 5, 30, and 50, and
3. the intra-class correlation coefficients (ICC) is set at 0.1, 0.2, and 0.3.

The number of groups is chosen so that the highest number should be sufficient. In practice, 50 groups is a frequently occurring number in organizational and school research, and 30 is the smallest acceptable number according to Maas & Hox (2005). Similarly, the group sizes are chosen so that the highest number should be sufficient. A group size of 30 is normal in educational research, and a group size of 5 is normal in family research and longitudinal research. The three ICCs included in the study span the customary range of ICC found in social, behavioral, and education sciences

(Gulliford, Ukoumunne, & Chinn, 1999).

The individual and group explanatory variables X_{ij} and Z_j are generated from the standard normal distribution. We set the fixed effect regression coefficients for all simulated models as follows: $\gamma_{00} = -1.0$, $\gamma_{01} = 0.3$, $\gamma_{10} = 0.3$, and $\gamma_{11} = 0.3$. The values of the fixed effect regression parameters do not change between models, while the variance σ_{u0}^2 varies according to the intraclass correlation coefficient, through the relation

$$\sigma_{u0}^2 = \frac{\frac{\pi^2}{3}\rho}{1-\rho}. \quad (3.2)$$

That is, σ_{u0} follows from the ICC and σ_e^2 . The variance of the random slope, σ_{u1}^2 , is set at 1 in all simulations. To simplify the simulation model, without loss of generality, the group random components u_{0j} and u_{1j} are assumed to be independent normal variables.

Among the three most common approximation methods for maximum likelihood estimation in a multilevel logistic regression model, the Laplace approximation technique is used in this simulation study. The reason behind choosing Laplace approximation instead of adaptive Gaussian quadrature is as follows. The adaptive Gaussian quadrature (AGQ) method produces good approximation when the dimension of random effects is 1, however, it starts giving worse results as the dimension of random effects increases. AGQ method is inaccurate if the dimension of random effects is greater than two. This is due to the limitation of the method in factorizing high-dimensional integrals into some integrals with low-dimension. Besides this, In a recent simulation study, Handayani et al. (2017) found that the Laplace approximation has the best performance (compared to AGQ and PQL) for obtaining maximum likelihood estimates in generalized linear mixed models in terms of their relative biases and mean square errors. Since in our simulation model the dimension of random effects is 2, we decide to use the Laplace approximation technique.

There are $3 \times 3 \times 3 = 27$ conditions. For each condition, we generated 1,000 simulated data sets. The **R** software (R version 3.6.0) is used for generating the simulated data sets and to estimate the model. Specifically, for the estimation we use the `glmer()` function from the **lme4** package. The absolute value for parameter convergence criterion was set at $1e^{-7}$ and the maximum number of function evaluations was set at 20000.

3.2 Analysis techniques

3.2.1 Percentage relative bias

The accuracy of the parameter estimates (fixed effects coefficients and residual variances) is quantified by the percentage relative bias. Let $\hat{\theta}$ stand for the estimate of the population parameter θ , then $\frac{\hat{\theta}-\theta}{\theta} \times 100$ indicates the percentage relative bias for parameter θ .

3.2.2 Non-coverage rate of confidence interval

The accuracy of the confidence interval is assessed by analyzing the observed non-coverage rate of the 95% confidence interval. For each estimated parameter in each simulated dataset, we calculate a non-coverage indicator, which is equal to 0 when the true value of that parameter lies inside the estimated 95% confidence interval (CI) and 1 when the true value lies outside this interval. Counting how many times the non-coverage indicator is equal to 1 in each simulation study, we can calculate the overall non-coverage rate and then compare it with the nominal rate of 5%.

Chapter 4

The simulation results

4.1 Rate of model convergence

The overall rate of model convergence varied from 59% to 100%. Logistic regression was used to investigate the impact of ICC, number of groups and group size on the convergence. The rate of convergence (percent converged) significantly improved with either an increase in the number of groups or an increase in the group size. The overall rate of convergence for number of groups 30, 50, and 100 was 91.76%, 94.87% and 97.6%, respectively. For groups of sizes 5, 30, and 50, the rate of convergence was 85.08%, 99.44%, and 99.7% . For the three ICC conditions of 0.1, 0.2, and 0.3 the rate of convergence was 92.1%, 95.91%, and 96.21%.

To explore the non-convergent samples Moineddin et al. (2007) examined 168 non-convergent simulated data sets with 30 groups and group size of 5. After investigation, it was concluded that non-convergence results from lack of sufficient variation in both the intercept and slope.

4.2 The bias in point estimates

Table 4.1 shows the percentage relative bias of estimates for different number of groups, group size, and ICC. We found that the relative biases for the estimates of standard errors of group-level variances are comparatively bigger than the fixed effect parameter estimates. Table 4.1 reveals that all the elements of the last two columns are negative numbers. That means the variance components are always underestimated. This problem of underestimation has been noted previously in simulation studies of Maas

& Hox (2005) and Moineddin et al. (2007). However, the biases for the estimates of standard errors of group-level residuals are less than 10% for number of groups 30 and group size 30, and are less than 5% when the size of the group was 50 and there were 50 groups. The biases for $ICC = 0.1$ are relatively higher than the ICC values 0.2 and 0.3.

From the results of percentage relative bias in Table 4.1, our main findings are:

- The estimates of the regression coefficients are always unbiased, even with small sample sizes, except some cases on level-2 variable coefficients.
- The variance components are always underestimated, even with 100 groups and 50 individuals. However, the bias decreases with the increase in the number of groups, and with number of groups 50 and group size 50 this bias is less than 5%.

Overall, results from our simulation studies regarding the accuracy of the point estimates lead to similar conclusions as to the previous studies, in particular to those of Moineddin et al. (2007) and Paccagnella (2011).

Table 4.1: The effect of number of groups, group size, and ICC on the percentage relative bias ($\frac{\hat{\theta}-\theta}{\theta} \times 100$) of the estimates.

| N. of groups | Group size | % | | | | | | | |
|-----------------|---------------|-----|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | ICC | Conv. | γ_{00} | γ_{10} | γ_{01} | γ_{11} | σ_{u0} | σ_{u1} |
| 30 | 5 | 0.1 | 69 | 3.54 | 5.16 | 11.71 | 3.77 | -11.33 | -16.64 |
| | | 0.2 | 80 | 2.09 | -3.52 | 5.24 | 7.18 | -18.02 | -16.06 |
| | | 0.3 | 82 | 2.25 | 7.4 | 4.78 | 8.81 | -14.14 | -12.12 |
| | 30 | 0.1 | 99 | -0.78 | 5.36 | 3.1 | 5.56 | -10.22 | -5.8 |
| | | 0.2 | 99 | -0.28 | -2.43 | 2.55 | 4.68 | -5.58 | -5.7 |
| | | 0.3 | 98 | 0.67 | 4.3 | -2.96 | 2.07 | -6.34 | -6.31 |
| | 50 | 0.1 | 100 | -1.22 | 0.63 | 1.62 | -0.75 | -6.6 | -5.14 |
| | | 0.2 | 99 | -0.56 | -1.31 | 1.08 | 0.97 | -5.51 | -5.15 |
| | | 0.3 | 100 | -0.26 | 0 | -1.59 | 2.8 | -5.67 | -6.34 |
| 50 | 5 | 0.1 | 76 | 0.98 | 0.4 | 2.97 | -6.18 | -16.21 | -14.61 |
| | | 0.2 | 89 | 1.8 | 3 | 5.15 | 6.02 | -14.21 | -16.44 |
| | | 0.3 | 91 | 0.12 | -1.49 | -1.6 | 5.03 | -12.63 | -17.28 |
| | 30 | 0.1 | 100 | -0.55 | 0.41 | -2.06 | -0.04 | -6.42 | -4.06 |

| N. of groups | Group size | ICC | % | | | | | | |
|-----------------|---------------|-----|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | | Conv. | γ_{00} | γ_{10} | γ_{01} | γ_{11} | σ_{u0} | σ_{u1} |
| 100 | 50 | 0.2 | 100 | -0.76 | -0.58 | 0.23 | -1.6 | -4.97 | -3.97 |
| | | 0.3 | 99 | -0.18 | -0.94 | 3.35 | -1.64 | -3.45 | -4.8 |
| | | 0.1 | 100 | 0.33 | 0.11 | 2.23 | 2.61 | -4.91 | -3.05 |
| | | 0.2 | 100 | -0.13 | 2.58 | 2.08 | -0.18 | -3.78 | -3.38 |
| | | 0.3 | 99 | 0.27 | -0.69 | 2.48 | 4.26 | -2.66 | -3.27 |
| | | 0.1 | 85 | -0.69 | 1.22 | -2.4 | -2.58 | -16.36 | -13.98 |
| | 5 | 0.2 | 96 | -1.04 | -0.46 | 2.23 | -0.69 | -14.46 | -14.9 |
| | | 0.3 | 97 | -1.83 | -2.36 | -0.07 | -1.85 | -11.74 | -18.15 |
| | | 0.1 | 100 | -0.11 | -1.44 | 0.22 | 2.14 | -4.27 | -2.72 |
| | | 0.2 | 100 | 0.04 | 1.41 | -0.21 | 2.21 | -2.65 | -2.59 |
| | | 0.3 | 100 | 0.58 | 0.23 | -0.13 | -1.03 | -2.7 | -3.32 |
| | | 0.1 | 100 | -0.01 | -0.51 | 1.04 | -0.4 | -2.85 | -2.33 |
| | 30 | 0.2 | 100 | -0.07 | 2.38 | 0.47 | -2.88 | -1.93 | -1.44 |
| | | 0.3 | 100 | 0.96 | 0.62 | -1.49 | 0.77 | -2.04 | -1.91 |

4.3 Non-coverage of confidence interval

Table 4.2, 4.3 and 4.4 display the average non-coverage rates of 95% profile confidence interval by number of groups, group size and ICC, respectively. Note that, we computed non-coverage rates on the standard deviation scale of variance components.

As we can see from Table 4.2, non-coverage rates for fixed effect parameters range from 5.9% to 6.8%. For the standard errors of group-level variances, non-coverage rates range from 6.3% to 7.9%. Overall, as the number of groups increases, the non-coverage rates of 95% profile likelihood interval for both the fixed effect and random effect parameters approach to the nominal 5%. Table 4.3 shows that as the group size increases, the non-coverage rates approach to nominal value 5%. Table 4.4 reveals no pattern of non-coverage rate for the change in ICC.

Table 4.5 shows the rates of non-coverage for each parameter in each simulated data set. Figure 4.1 depicts how the non-coverage rates are scattered around the nominal rate of 0.5 (the horizontal red line) by a dot plot. The main findings from non-coverage rates are:

- The mean non-coverage rate for profile CI is 6.39%, and non-coverage rates are more close to 6%, than the nominal value 5% (as Figure 4.1 reveals).
- In some cases, profile likelihood confidence interval performs poorly for the variance components (e.g., the three cases for 100 groups with 5 individuals per group). This suggests, there might be other factors that influence the coverage rate of profile likelihood confidence interval.

Table 4.2: Non-coverage of the 95% confidence interval by number of groups.

| Parameters | N. of groups = 30 | N. of groups = 50 | N. of groups = 100 |
|---------------|-------------------|-------------------|--------------------|
| γ_{00} | 0.064 | 0.060 | 0.059 |
| γ_{10} | 0.068 | 0.059 | 0.057 |
| γ_{01} | 0.065 | 0.059 | 0.062 |
| γ_{11} | 0.067 | 0.061 | 0.058 |
| σ_{u0} | 0.070 | 0.064 | 0.067 |
| σ_{u1} | 0.067 | 0.063 | 0.079 |

Table 4.3: Non-coverage of the 95% confidence interval by group size.

| Parameters | Group size = 5 | Group size = 30 | Group size = 50 |
|---------------|----------------|-----------------|-----------------|
| γ_{00} | 0.066 | 0.059 | 0.059 |
| γ_{10} | 0.069 | 0.061 | 0.054 |
| γ_{01} | 0.065 | 0.062 | 0.059 |
| γ_{11} | 0.064 | 0.057 | 0.066 |
| σ_{u0} | 0.064 | 0.073 | 0.064 |
| σ_{u1} | 0.079 | 0.068 | 0.061 |

Table 4.4: Non-coverage of the 95% confidence interval by Intra-class correlation (ICC).

| Parameters | ICC = 0.1 | ICC = 0.2 | ICC = 0.3 |
|---------------|-----------|-----------|-----------|
| γ_{00} | 0.064 | 0.059 | 0.060 |
| γ_{10} | 0.060 | 0.061 | 0.063 |

| Parameters | ICC = 0.1 | ICC = 0.2 | ICC = 0.3 |
|---------------|-----------|-----------|-----------|
| γ_{01} | 0.063 | 0.060 | 0.064 |
| γ_{11} | 0.061 | 0.064 | 0.062 |
| σ_{u0} | 0.056 | 0.073 | 0.072 |
| σ_{u1} | 0.070 | 0.069 | 0.070 |

Table 4.5: The effect of number of groups, group size, and ICC on the non-coverage of the 95% profile confidence interval.

| Number of groups | Group size | ICC | γ_{00} | γ_{10} | γ_{01} | γ_{11} | σ_{u0} | σ_{u1} |
|---------------------|---------------|-----|---------------|---------------|---------------|---------------|---------------|---------------|
| 30 | 5 | 0.1 | 0.053 | 0.089 | 0.065 | 0.063 | 0.019 | 0.058 |
| | | 0.2 | 0.074 | 0.049 | 0.07 | 0.069 | 0.069 | 0.064 |
| | | 0.3 | 0.07 | 0.082 | 0.075 | 0.062 | 0.076 | 0.053 |
| | 30 | 0.1 | 0.072 | 0.062 | 0.063 | 0.072 | 0.081 | 0.072 |
| | | 0.2 | 0.052 | 0.068 | 0.058 | 0.056 | 0.091 | 0.074 |
| | | 0.3 | 0.057 | 0.074 | 0.068 | 0.063 | 0.078 | 0.089 |
| | 50 | 0.1 | 0.061 | 0.059 | 0.066 | 0.078 | 0.059 | 0.056 |
| | | 0.2 | 0.059 | 0.065 | 0.059 | 0.083 | 0.075 | 0.061 |
| | | 0.3 | 0.074 | 0.062 | 0.061 | 0.057 | 0.081 | 0.074 |
| 50 | 5 | 0.1 | 0.073 | 0.061 | 0.055 | 0.06 | 0.027 | 0.082 |
| | | 0.2 | 0.061 | 0.065 | 0.052 | 0.066 | 0.052 | 0.064 |
| | | 0.3 | 0.073 | 0.069 | 0.067 | 0.072 | 0.09 | 0.055 |
| | 30 | 0.1 | 0.057 | 0.054 | 0.065 | 0.065 | 0.067 | 0.057 |
| | | 0.2 | 0.047 | 0.056 | 0.065 | 0.059 | 0.078 | 0.064 |
| | | 0.3 | 0.054 | 0.073 | 0.056 | 0.049 | 0.059 | 0.058 |
| | 50 | 0.1 | 0.067 | 0.053 | 0.065 | 0.043 | 0.075 | 0.064 |
| | | 0.2 | 0.056 | 0.058 | 0.05 | 0.069 | 0.062 | 0.059 |
| | | 0.3 | 0.052 | 0.044 | 0.059 | 0.065 | 0.063 | 0.062 |
| 100 | 5 | 0.1 | 0.069 | 0.061 | 0.069 | 0.06 | 0.038 | 0.113 |
| | | 0.2 | 0.06 | 0.074 | 0.057 | 0.055 | 0.102 | 0.111 |
| | | 0.3 | 0.061 | 0.072 | 0.072 | 0.064 | 0.103 | 0.112 |
| | 30 | 0.1 | 0.067 | 0.047 | 0.057 | 0.044 | 0.074 | 0.071 |
| | | 0.2 | 0.057 | 0.065 | 0.066 | 0.043 | 0.074 | 0.067 |

| Number of groups | Group size | ICC | γ_{00} | γ_{10} | γ_{01} | γ_{11} | σ_{u0} | σ_{u1} |
|---------------------|---------------|-----|---------------|---------------|---------------|---------------|---------------|---------------|
| | | 0.3 | 0.063 | 0.049 | 0.063 | 0.063 | 0.055 | 0.062 |
| | 50 | 0.1 | 0.055 | 0.054 | 0.063 | 0.061 | 0.062 | 0.054 |
| | | 0.2 | 0.063 | 0.05 | 0.06 | 0.076 | 0.056 | 0.055 |
| | | 0.3 | 0.039 | 0.043 | 0.05 | 0.059 | 0.041 | 0.065 |

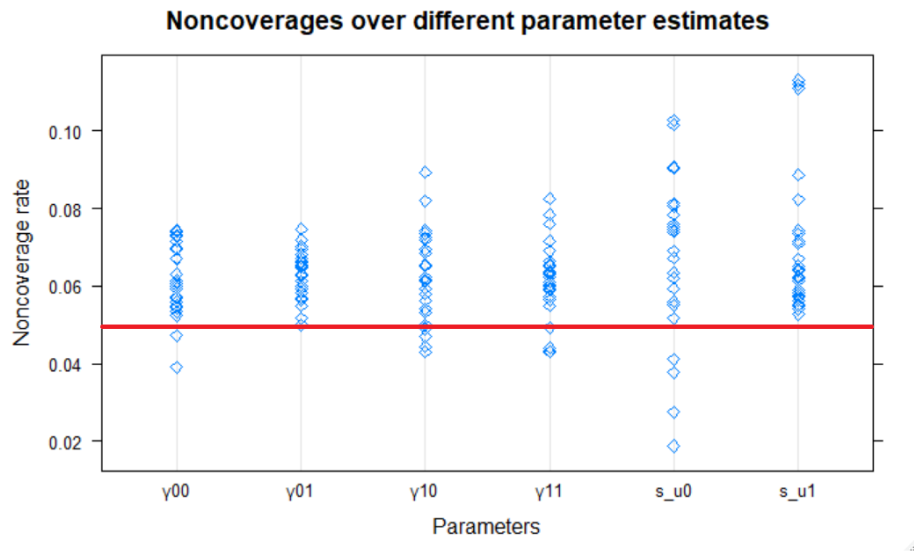


Figure 4.1: Dot plots of non-coverage rates of the profile likelihood confidence interval over different parameter estimates.

Chapter 5

Conclusion

In this study, we perform Monte Carlo simulation studies to investigate the sample behavior of estimates in a binary response multilevel model using the Laplace approximation estimation procedure. Particularly, we analyze the influence of different factors (different number of groups, different group sizes, different intra-class correlation coefficients) on the accuracy of parameter estimates (fixed effects and variance components) and their profile likelihood confidence intervals.

Despite the limitations of simulation studies, our findings can offer some suggestions for sample size selection in multilevel logistic regression. With a group of 50 and group size 30, we found that the non-coverage rates for fixed effect parameters range from 4.7% to 7.3% and for level-2 variance components from 5.7% to 7.8%. This is clearly different from the nominal 5%, however, in practice probably acceptable (Maas & Hox, 2005). Besides, with this sample size, the model convergence rate is almost 100%, and the bias of variance components are less 6%. Thus, based on our findings, we recommend a minimum of 50 groups with at least 30 units per group.

In our study, we assumed group level residuals are normally distributed. This may not be the case always. Our results are comparable with the simulation results reported by Maas & Hox (2005), Moineddin et al. (2007), and Paccagnella (2011), when only those conditions are considered that are similar to the conditions in our simulations. The results of some simulation studies, reviewed in this project, may appear somewhat inconsistent. However, this is probably the result of using different simulation designs and different simulated conditions.

The main findings from this research can be summarized as follows:

- The estimates of the regression coefficients are always unbiased, even with small

sample sizes, except some cases on level-2 variable coefficients.

- The variance components are always underestimated, even with 100 groups and 50 individuals. However, the bias decreases with the increase in number of groups, and with number of groups 50 and group size 50 this bias is less than 5%.
- The mean non-coverage rate for profile likelihood confidence interval is 6.39%, and non-coverage rates are more close to 6%, than the nominal value 5%.
- In some cases, profile likelihood confidence interval performs poorly for the variance components (e.g., the three cases for 100 groups with 5 individuals per group). This suggests, there might be other factors that influence the coverage rate of profile likelihood confidence interval.

Appendix A

R codes used in this study

```
library(lme4)

# Setting the parameters values
g_00 <- -1
g_01 <- 0.3
g_10 <- 0.3
g_11 <- 0.3

s2_1 <- 1                      # Variance standardized to 1
s01 <- 0                      # Covariance assumed zero
s2_e <- (pi^2)/3              # Variance of the logistic distribution

# Function for simulating data from two-level logistic regression model

sim_fun <- function(J, n_j, ICC){
  s2_0 <- s2_e / (1/ICC-1)    # Variance corresponding to specific ICC

  n_j <- rep(n_j, J)         # Number of individuals in jth group
  N <- sum(n_j)              # Number of groups

  # Simulate the covariate values for this sample size.

  z <- rnorm(J)
  x <- rnorm(N)
```

```

# Generate (u_0j, u_1j) from a bivariate normal

mu <- c(0, 0)
sig <- matrix(c(s2_0, s01, s01, s2_1), ncol = 2)
u <- MASS::mvrnorm(J, mu = mu, Sigma = sig)

# Now form the linear predictor. First, the group-level predictor.
# There are J of these.

pi_0 <- g_00 + g_01*z + as.vector(u[,1])
pi_1 <- g_10 + g_11*z + as.vector(u[,2])
eta <- rep(pi_0, n_j) + rep(pi_1, n_j) * x

# Transform back to the probability scale.

p <- exp(eta) / (1+exp(eta))

# Simulate a Bernoulli from each individual distribution.

y <- rbinom(N, 1, p)

sim_data <- data.frame(Y=y, X=x, Z=rep(z,n_j), group=rep(1:J,n_j))
}
# End of the function 'sim_fun'

# Function for calculating percentage relative bias of the parameter estimates
# and calculating non-coverage rates.

analysis <- function(J, n_j, ICC) {
  sim_dat <- sim_fun(J, n_j, ICC)
  fit <- glmer(Y ~ X + Z + X:Z + (X||group), family=binomial, data=sim_dat,
              control=glmerControl(optCtrl=list(maxfun=2e5)))

  s2_0 <- s2_e / (1/ICC-1) ##variance corresponding to specific ICC

  true_val <- c(g_00, g_10, g_01, g_11, sqrt(s2_0), 1)
  theta_hat <- c(as.numeric(fixef(fit)), as.data.frame(VarCorr(fit))[,5])

```

```

perc_rel_bias <- ((theta_hat-true_val)/true_val)*100

ci_profile <- confint.merMod(object = fit, method = "profile", quiet = TRUE)

nc_u0 <- ifelse (ci_profile[1] < sqrt(s2_0) & ci_profile[7] > sqrt(s2_0), 0, 1)
nc_u1 <- ifelse (ci_profile[2] < sqrt(s2_1) & ci_profile[8] > sqrt(s2_1), 0, 1)
nc_int <- ifelse (ci_profile[3] < g_00          & ci_profile[9] > g_00,          0, 1)
nc_x  <- ifelse (ci_profile[4] < g_10          & ci_profile[10] > g_10,          0, 1)
nc_z  <- ifelse (ci_profile[5] < g_01          & ci_profile[11] > g_01,          0, 1)
nc_xz <- ifelse (ci_profile[6] < g_11          & ci_profile[12] > g_11,          0, 1)

noncoverage <- c(nc_int, nc_x, nc_z, nc_xz, nc_u0, nc_u1)
return (c(perc_rel_bias, noncoverage))
}
# End of the function 'analysis'

# Combination 01:
c01 <- replicate(1000, tryCatch(analysis(30, 5, 0.1), error=function(err) rep(NA,12)))
r <- rowMeans(apply(c01, 2, unlist), na.rm = TRUE)
perc_conv <- 1 - sum(is.na(c01[1,]))/1000

# Combination 02:
c02 <- replicate(1000, tryCatch(analysis(30, 5, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c02, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c02[1,]))/1000)

# Combination 03:
c03 <- replicate(1000, tryCatch(analysis(30, 5, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c03, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c03[1,]))/1000)

# Combination 04:
c04 <- replicate(1000, tryCatch(analysis(30, 30, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c04, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c04[1,]))/1000)

# Combination 05:

```

```
c05 <- replicate(1000, tryCatch(analysis(30, 30, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c05, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c05[1,]))/1000)

# Combination 06:
c06 <- replicate(1000, tryCatch(analysis(30, 30, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c06, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c06[1,]))/1000)

# Combination 07:
c07 <- replicate(1000, tryCatch(analysis(30, 50, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c07, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c07[1,]))/1000)

# Combination 08:
c08 <- replicate(1000, tryCatch(analysis(30, 50, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c08, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c08[1,]))/1000)

# Combination 09:
c09 <- replicate(1000, tryCatch(analysis(30, 50, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c09, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c09[1,]))/1000)

# Combination 10:
c10 <- replicate(1000, tryCatch(analysis(50, 5, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c10, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c10[1,]))/1000)

# Combination 11:
c11 <- replicate(1000, tryCatch(analysis(50, 5, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c11, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c11[1,]))/1000)

# Combination 12:
c12 <- replicate(1000, tryCatch(analysis(50, 5, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c12, 2, unlist), na.rm = TRUE))
```

```
perc_conv <- c(perc_conv, 1 - sum(is.na(c12[1,]))/1000)

# Combination 13:
c13 <- replicate(1000, tryCatch(analysis(50, 30, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c13, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c13[1,]))/1000)

# Combination 14:
c14 <- replicate(1000, tryCatch(analysis(50, 30, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c14, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c14[1,]))/1000)

# Combination 15:
c15 <- replicate(1000, tryCatch(analysis(50, 30, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c15, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c15[1,]))/1000)

# Combination 16:
c16 <- replicate(1000, tryCatch(analysis(50, 50, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c16, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c16[1,]))/1000)

# Combination 17:
c17 <- replicate(1000, tryCatch(analysis(50, 50, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c17, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c17[1,]))/1000)

# Combination 18:
c18 <- replicate(1000, tryCatch(analysis(50, 50, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c18, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c18[1,]))/1000)

# Combination 19:
c19 <- replicate(1000, tryCatch(analysis(100, 5, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c19, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c19[1,]))/1000)
```

Combination 20:

```
c20 <- replicate(1000, tryCatch(analysis(100, 5, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c20, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c20[1,]))/1000)
```

Combination 21:

```
c21 <- replicate(1000, tryCatch(analysis(100, 5, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c21, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c21[1,]))/1000)
```

Combination 22:

```
c22 <- replicate(1000, tryCatch(analysis(100, 30, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c22, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c22[1,]))/1000)
```

Combination 23:

```
c23 <- replicate(1000, tryCatch(analysis(100, 30, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c23, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c23[1,]))/1000)
```

Combination 24:

```
c24 <- replicate(1000, tryCatch(analysis(100, 30, 0.3), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c24, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c24[1,]))/1000)
```

Combination 25:

```
c25 <- replicate(1000, tryCatch(analysis(100, 50, 0.1), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c25, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c25[1,]))/1000)
```

Combination 26:

```
c26 <- replicate(1000, tryCatch(analysis(100, 50, 0.2), error=function(err) rep(NA,12)))
r <- rbind(r, rowMeans(apply(c26, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c26[1,]))/1000)
```

Combination 27:

```
c27 <- replicate(1000, tryCatch(analysis(100, 50, 0.3), error=function(err) rep(NA,12)))
```

```
r <- rbind(r, rowMeans(apply(c27, 2, unlist), na.rm = TRUE))
perc_conv <- c(perc_conv, 1 - sum(is.na(c27[1,]))/1000)

r[,1:6]      # Relative bias of the six estimates
r[,7:12]     # Non-coverage rates of profile CI of the six estimates
perc_conv*100 # Percentages of Convergence in the 27 conditions
```

References

- Akter, N. J., & Khan, M. H. R. (2018). Effect of sample size on the profile likelihood estimates for two-stage hierarchical linear models. *Journal of Biomedical Analytics*, 1(2), 81–89.
- Austin, P. C. (2005). Bias in penalized quasi-likelihood estimation in random effects logistic regression models when the random effects are not normally distributed. *Communications in Statistics-Simulation and Computation*, 34(3), 549–565.
- Bates, D. M. (2010). Lme4: Mixed-effects modeling with r.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9–25.
- Breslow, N. E., & Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82(1), 81–91.
- Clarke, P., & Wheaton, B. (2007). Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods. *Sociological Methods & Research*, 35(3), 311–351.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Wiley.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London, UK: Arnold Publishers.
- Goldstein, H., & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(3), 505–513.
- Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: Data from the health survey for england 1994. *American Journal of Epidemiology*, 149(9), 876–883.

- Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, 26(1), 441–462.
- Handayani, D., Notodiputro, K. A., Sadik, K., & Kurnia, A. (2017). A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (glmm). In *AIP conference proceedings* (Vol. 1827, p. 020033). AIP Publishing.
- Hox, J. (1998). Multilevel modeling: When and why. In *Classification, data analysis, and data highways* (pp. 147–154). Springer.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications, second edition (quantitative methodology series)*. Routledge. Retrieved from <https://www.xarg.org/ref/a/1848728468/>
- Kreft, I. G. G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Unpublished manuscript, California State University at Los Angeles. Retrieved July 6, 2005 from <http://www.calstatela.edu/faculty/ikreft/quarterly.html>.
- Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: An example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3), 325–335.
- Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58(2), 127–137.
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92.
- McNeish, D. (2016). Estimation methods for mixed logistic models with few clusters. *Multivariate Behavioral Research*, 51(6), 790–804.
- Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, 7(1), 34.
- Mok, M. (1995). *Sample size requirements for 2-level designs in educational research*.
- Paccagnella, O. (2011). Sample size and accuracy of estimates in multilevel models. *Methodology*.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace

- approximation. *Journal of Computational and Graphical Statistics*, 9(1), 141–157.
- Sastry, N., Ghosh-Dastidar, B., Adams, J., & Pebley, A. R. (2006). The design of a multilevel survey of children, families, and communities: The los angeles family and neighborhood survey. *Social Science Research*, 35(4), 1000–1024.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE Publications.
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using stata, r, mplus, and spss. *International Review of Social Psychology*, 30(1).
- Venzon, D. J., & Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, 37(1), 87. <http://doi.org/10.2307/2347496>