2009

Association Models for Clustered Data with Binary and Continuous Responses

Lanjia Lin

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

ASSOCIATION MODELS FOR CLUSTERED DATA WITH BINARY AND CONTINUOUS RESPONSES

By

LANJIA LIN

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2009

The members of the Committee approve the Dissertation of Lanjia Lin defended on April 8, 2009.

Debajyoti Sinha
Professor Directing Dissertation

Myra Hurt
Outside Committee Member

Stuart R. Lipsitz
Committee Member

Daniel McGee
Committee Member

Approved:

Daniel McGee, Chair, Department of Statistics

Joseph Travis, Dean, College of Arts and Sciences

The Graduate School has verified and approved the above named committee members.

This dissertation is dedicated to my husband Peng and my parents Shaokang and Yonglan.

ACKNOWLEDGEMENTS

My first acknowledgment must go to my advisor, Dr. D. Sinha. I would like to thank him for his invaluable guidance on my dissertation research, and his patience, his encouragement and his support during my graduate career at FSU. I would also like to thank Dr. S.R. Lipsitz from Harvard medical school and Dr. D. Bandyopadhya from Medical University of South Carolina for their suggestions and help on my research work. My other committee members, Dr. D. McGee and Dr. M. Hurt, also have my gratitude.

I am also very grateful to all the faculty, especially Dr. X. Niu and Dr. F. Huffer, and all the staff in the Department of Statistics for their generous help during my time at FSU.

A very big and sincere thanks goes to my husband Peng, for his unconditional support on my graduate study. I am also indebted to my dearest parents, Shaokang and Yonglan, for their selfless love and emotional support.

— Lanjia

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation develops novel single random effect models as well as bivariate correlated random effects model for clustered data with bivariate mixed responses. Logit and identity link functions are used for the binary and continuous responses. For the ease of interpretation of the regression effects, random effect of the binary response has bridge distribution so that the marginal model of mean of the binary response after integrating out the random effect preserves logistic form. And the marginal regression function of the continuous response preserves linear form. Within-cluster and within-subject associations could be measured by our proposed models. For the bivariate correlated random effects model, we illustrate how different levels of the association between two random effects induce different Kendall's τ values for association between the binary and continuous responses from the same cluster. Fully parametric and semi-parametric Bayesian methods as well as maximum likelihood method are illustrated for model analysis. In the semiparametric Bayesian model, normality assumption of the regression error for the continuous response is relaxed by using a nonparametric Dirichlet Process prior. Robustness of the bivariate correlated random effects model using ML method to misspecifications of regression function as well as random effect distribution is investigated by simulation studies. The Bayesian and likelihood methods are applied to a developmental toxicity study of ethylene glycol in mice.

CHAPTER 1

INTRODUCTION

1.1 Clustered Data with Mixed Responses

In many biomedical studies, the basic sampling unit is a cluster of subjects, such as siblings in a family or rats in a litter. Further, often each member of the cluster may have multiple outcomes, including both discrete and continuous endpoints (called mixed responses). There has been some recent interests in modeling and analysis of mixed responses data with discrete and continuous outcomes (e.g., Cox and Wermuth [1]; Sammel, Ryan, and Legler [2]; Dunson [3]; de Leon and Carriere [4]). The development toxicity study of industrial chemical Ethylene Glycol (EG) [5] conducted through the National Toxicity Program (NTP) is an example. In this study, one of four different dose levels of Ethylene Glycol (EG), a high-volume industrial chemical, was administered to 94 pregnant mice over the period of major organogenesis, beginning just after implantation. Both binary and continuous responses, viz. (a) fetal malformation and (b) fetal weight, of 1028 fetuses from 94 litters (clusters) were measured.

To analyze clustered data, random effect is usually used in mixed model to characterize within-cluster associations. When the bivariate responses are both continuous and approximately Gaussian, the traditional linear model framework can be used for inference. However, when the response is not assumed to be normally distributed and the mean is not in linear form, generalized linear model is often used for data analysis. The presence of mixed endpoints (categorical/ordinal as well as continuous outcomes), as arising in the data from the development toxicity study of EG, complicates the situation due to the lack of a discrete multivariate analog of the multivariate Gaussian distribution [6]. Also analysis of such kind of data needs to take care of the associations among responses within the same cluster, as well as the associations between the mixed responses within the same subject.

1.2 Existing Models

Analysis for the clustered bivariate mixed data with discrete and continuous responses has been proposed by many authors including Fitzmaurice and Laird [6], Catalano and Ryan [7], Regan and Catalano [8], Geys, et al. [9], Gueorguieva and Agresti [10], etc.

Currently available models for this type of data are of three types. In some models, the joint distribution of bivariate outcomes was formed by product of the marginal density function of the continuous response and the conditional density of binary response given the continuous response. For example, Catalano and Ryan [7], henceforth CR, suggested a model assuming that the discrete outcome is a discretization of an unobserved continuous latent variable, and this latent variable and the continuous response within same subject have say, a bivariate normal distribution. Joint distribution of the two responses is then formed by a product of the marginal distribution of the continuous response and the conditional distribution of the binary response conditioned on the continuous response. Linear and probit link functions are assigned to the continuous response and the binary response, respectively.

Similarly there are models based on specifying the joint distribution as the marginal distribution of the binary response multiplied by the conditional distribution of continuous response given the binary response. Fitzmaurice and Laird [6] proposed a model with logit link function for marginal probability of binary response and a normal distribution for the continuous response given the binary response. Cox and Wermuth [1] compared different joint distribution models for analyzing data with quantitative and qualitative responses. Parameter estimates of the above models are obtained using either generalized estimating equations (GEE) method or the maximum likelihood method via expectation-maximization (EM) algorithm under different types of models.

In the third modeling approach, Dunson [3] described a Bayesian approach using a mixture of generalized linear models for joint distribution of latent variables for the clustered mixed outcomes. Cluster and subject level latent variables were assigned multivariate normal densities or linked to variables with simple exponential families. Markov chain Monte Carlo (MCMC) algorithms are also developed in [3] for posterior distribution estimations of parameters.

Irrespective of modeling strategies, the methods based on GEE, e.g. Fitzmaurice and Laird [6], treat the associations within clusters and within subjects as nuisance. On the

contrary, the likelihood and Bayesian methods focus full model estimation of the association structures (e.g. Dunson [3]). One drawback of the models proposed by Catalano and Ryan [7] and Dusnson [3] is that regression parameters for the binary response using the probit link function do not have an odds ratio interpretation. The model proposed by Fitzmaurice and Laird [6] (FL) have the attractive feature that the marginal regression parameters of both responses have good physical interpretations. However, in FL model, the specification of the regression model conditional on the cluster-specific random effects is completely unknown. In clustered data, the magnitude of the marginal regression parameters, particularly for the binary response, gets attenuated from that of the conditional regression parameters due to the heterogeneity of clusters. This degree of attenuation is important for future study design as well as understanding the expected effect of covariates on other populations and on individual subjects.

1.3 Motivation and Purpose

To avoid the above shortcomings of CR and FL modeling, we propose novel random effect models with practical physical interpretations of the marginal as well as conditional (given the cluster-specific random effects) regression functions associated with both discrete and continuous responses. We are focused on joint modeling the regression effects on the binary and continuous outcomes, as well as associations between the mixed outcomes from the same subject and from different subjects within the same cluster. Like FL model, we assume logit link function and identity link function for the binary outcome and continuous outcome, respectively. Both single random effect models and correlated bivariate random effects model are proposed and compared in the following chapters.

It is known that the marginal regression model of mean of the binary response (after integrating out the random effects) usually does not preserve logistic form when the (subject-specific) model conditional on the random effects is logistic. The distribution of the random effects in our subject-specific model for the binary response has a Bridge distribution of Wang and Louis [11]. This allows the marginal probability of the binary response, integrated over the random intercept, to have a logistic structure with an odds ratio interpretation of the marginal regression effect. The regression parameters in the marginal logistic regression model are proportional to the corresponding regression parameters in the subject-specific

conditional logistic model. Also, the marginal mean of the continuous response after integrating out random effects still has a linear form. In this way, our model has practical physical interpretations of the marginal as well as conditional regression functions associated with both responses.

Besides exploring flexible and adequately compact modeling strategy, we will also focus on developing implementable Bayesian methods for clustered mixed responses data in this article. Bayesian method has several advantages over the frequentist method, like appropriate inferences of parameters based on exact posterior distribution can be obtained and prior knowledge could be specified to bring extra information for more precision estimates of parameters [3].

In this dissertation, we use fully parametric as well as semiparametric Bayesian methods to analyze the specified joint models of mixed responses. In the parametric Bayesian method, we assign Bernoulli distribution to the binary response given the random effect, and the normal densities with a constant variance for the error term in the conditional model of the continuous response. In the mixed effects model, inferences are often sensitive to the form of the density of the regression errors of response. We extend the linear model of the continuous response to a large class of symmetric density via using semiparametric Bayesian approach. To avoid the restrictive normality assumption with constant variance for regression errors of the continuous response, we assume the regression errors to have normal distribution with mean zero and subject-specific variances. Moreover, subject-specific variance is assigned with an unknown distribution which has a nonparametric Dirichlet Process (DP) [12] prior. In this way, the normality assumption of continuous response is replaced by a large semiparametric class of symmetric and unimodal density [13].

In the CR and FL models, they considered the maximum likelihood (ML) estimation as quite complicated and intractable, and hence they proposed GEE method for computational convenience. Contrary to that, likelihood in our proposed correlated bivariate random effect model is amenable to maximization through routine nonadaptive Gaussian quadrature techniques using standard software SAS. Besides the Bayesian approaches, we also apply ML method to the correlated bivariate random effect model for data analysis. Robustness of this model to regression function misspecification as well as random effect distribution misspecification is investigated through a simulation study.

Rest chapters are organized as follows. In Chapter 2, we introduce two single random

effect models as well as a correlated bivariate random effects model for clustered data with both binary and continuous responses. In the bivariate random effects model, correlated distribution of two cluster-specific random effects, presented as separate random intercepts for two types of responses, is specified. Associations of the mixed responses within the same cluster are also discussed using Kendal's τ for bivariate random effects model. In Chapter 3, we discuss the parametric Bayesian method with prior elicitation and posterior computation for proposed models. We present parametric Bayesian methods for analysis of the motivation data from a developmental toxicity study in Chapter 4. This section also provides a model selection tool to choose the most appropriate model from the set of available models. Semiparametric Bayesian methods for analyzing the correlated random effects model as well as Dirichlet process (DP) prior with application is discussed in Chapter 5. The maximum likelihood method with application is illustrated in Chapter 6. Robustness of the correlated random effects model to model misspecification and random effect misspecification is also investigated. Conclusion and a discussion of future work is given in Chapter 7.

CHAPTER 2

ASSOCIATION MODELS

We review generalized linear model (GLM) and generalized linear mixed model (GLMM) at the beginning of this chapter. Two different single random effect models and a correlated bivariate random effects model are then proposed for joint modeling clustered data with binary and continuous responses, each of which is modeled by a generalized linear mixed model. Within-cluster and within-subject associations between the binary response and the continuous response are also investigated in this chapter.

2.1 Generalized Linear Mixed Model

In statistics, linear regression is widely used to model a response which is continuous and has a Gaussian distribution. Generalized linear models extend ordinary linear regression models when response of data is not assumed to be normally distributed and the mean is not in linear form. A generalized linear model consists of three components: a random component, a systematic component and a link function [14].

In the random component, a distribution from natural exponential family is given for response variable Y with independent observations $(y_1, ..., y_N)$:

$$f(y_i; \theta_i, \varphi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\varphi) + c(y_i, \varphi)\} \qquad (i = 1, ..., N),$$

where θ_i is the natural parameter and φ is the dispersion parameter.

The systematic component specifies a linear function for covariates $X_i = (x_{i1}, ..., x_{ip})$ of Y_i:

$$\eta_i = \Sigma_k x_{ik}\beta_k \qquad (i = 1, ..., N),$$

where x_{ik} is the predictor k $(k = 1, ..., p)$ for subject i.

The link function $g(\cdot)$ relates the $\mu_i = E(Y_i)$ to the systematic component by

$$\eta_i = g(\mu_i).$$

Parameters that characterize effects of covariates in generalized linear model are called fixed effects. However, subjects are often clustered in many real problems. The generalized linear mixed model allows random effects as well as fixed effects in the linear predictor function. Let y_{ij} denote observation j $(j = 1, ..., n_i)$ in cluster i, $\boldsymbol{u}_i$ denote the cluster-specific random effects for cluster i, $\boldsymbol{z}_{ij}$ be the corresponding explanatory variables, and $\mu_{ij} = E(Y_{ij}|\boldsymbol{u}_i)$. Conditioned on $\boldsymbol{u}_i$, y_{ij} are considered to be independent and exchangeable. However, marginally they are non-negatively correlated. The GLMM has the following form

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{u}_i,$$

where $\boldsymbol{x}_{ij}$ is the vector of covariates for subject j in cluster i, $\boldsymbol{\beta}$ is the vector of regression parameters for fixed effects.

In GLMM, the main interest is to investigate the fixed effects on response. Random effects are used to characterize the correlation between observations within the same cluster. Usually the random effect is univariate, and assumed to be normally distributed with mean 0 and variance σ_r. When $\sigma_r = 0$, the GLMM simplifies to GLM. Parameters in the random effects model are conditional effects while effects in the marginal models (integrated out random effects) are population averaged over all clusters. The marginal effects of explanatory variables are usually smaller than the effects conditioned on the random effects.

2.2 Cluster-specific Random Effect Model (Model 1)

In this section, we propose a model with single cluster-specific random effect to joint analyze the binary and continuous responses from clustered data.

For the jth $(j = 1, ..., n_i)$ subject from cluster i $(i = 1, ..., N)$, let Y_{1ij} be the binary response with value 1 for success and 0 otherwise, Y_{2ij} be the continuous response, and X_{ij} be the $K \times 1$ covariate vector for subject (i, j).

2.2.1 Model 1

Model 1 specifies a single cluster-specific random effect for both mixed responses, denoted as the random intercept R_i. We assume the binary response Y_{1ij} given R_i to be independent

Bernoulli distributed with success probability $p_{ij} = P(Y_{1ij} = 1|R_i, X_{ij})$, and the continuous response Y_{2ij} given the binary response Y_{1ij} and random effect R_i to have independent normal distribution with mean $E[Y_{2ij}|R_i, Y_{1ij}, X_{ij}]$ and constant variance. Logit link function and identity link function are assigned to Y_{1ij} and Y_{2ij}, respectively. Then we have:

$$\text{logit}(p_{ij}) = X_{ij}^T\boldsymbol{\alpha} + R_i \tag{2.1}$$

$$Y_{2ij} = X_{ij}^T\boldsymbol{\beta} + \gamma(Y_{1ij} - p_{1ij}) + \lambda R_i + \epsilon_{ij} \ , \tag{2.2}$$

where $p_{1ij} = E[Y_{1ij}|X_{ij}]$ is the marginal probability of $Y_{1ij} = 1$ after integrated out R_i, and ϵ_{ij} is normally distributed with mean 0 and variance σ^2.

From Model 1, the binary response Y_{1ij} and continuous response $Y_{2ij'}$ from different subjects within the same cluster share the same cluster-specific random effect R_i with coefficient parameter λ. Association between Y_{1ij} and Y_{2ij} within the same subject is induced by shared R_i as well as direct dependence of Y_{2ij} on Y_{1ij} measured by γ. Hence, this model insures that association between the mixed responses from the same subject is greater than the association between the mixed responses from different subjects within the same cluster.

2.2.2 Bridge Distribution

One drawback of the random effect models for clustered data is that the marginal regression model of the binary response Y_{1ij} (after integrating out the random effect R_i) does not preserve the logistic form as in the conditional model when R_i is assumed normally distributed. Hence, instead of normal distribution, we use a bridge distribution, proposed by Wang and Louis [11], as the marginal density of R_i so that parameters in the logistic regression model for Y_{1ij} could be marginally interpreted.

The bridge distribution is derived as follows [11]: for a general inverse link function $H(\cdot)$ which is assumed monotone, increasing and twice differentiable, let $F_\phi(r)$ denote the bridge distribution for cluster-specific random effect r with parameter ϕ. The bridge distribution for cluster random effect has the property that both the marginal and conditional link function have the same form, i.e. $F_\phi(\cdot)$ satisfies

$$\int H(X^T\boldsymbol{\alpha} + r)dF_\phi(r) = H(\phi X^T\boldsymbol{\alpha} + c), \tag{2.3}$$

where ϕ and c are unknown. When H is the cumulative distribution function of a symmetric distribution, $c = 0$. By differentiating and taking Fourier transforms of both sides of (2.3),

the density function of bridge distribution is

$$f_\phi(x) = \frac{1}{2\pi}\int e^{i(c/\phi - x)\xi}\frac{\mathcal{F}h(\xi/\phi)}{\mathcal{F}h(\xi)}d\xi,$$

where $\mathcal{F}$ is the Fourier transform operation defined as

$$\mathcal{F}(h(x)) = \mathcal{F}h(\xi) = \int e^{-i\xi x}h(x)dx,$$

and $h = H'$.

For logit link function, bridge distribution has symmetric density

$$f_\phi(x) = \frac{1}{2\pi}\frac{\sin(\phi\pi)}{\cosh(\phi x) + \cos(\phi\pi)} \qquad (0 < \phi < 1, -\infty < x < \infty) \tag{2.4}$$

with mean 0, variance $\pi^2(\phi^{-2} - 1)/3$, and the characteristic function

$$\varphi_\phi(x) = \frac{\sinh(\pi x)}{\phi\sinh(\pi x/\phi)}.$$

The bridge density for logit link has a slightly heavier tail and is more peaked than the normal density. It also has explicit cumulative distribution function (cdf) and inverse cdf forms:

$$F_\phi(x) = 1 - \frac{1}{\pi\phi}\left[\frac{\pi}{2} - \arctan\left\{\frac{\exp(\phi x) + \cos(\phi\pi)}{\sin(\phi\pi)}\right\}\right] \qquad (-\infty < x < \infty)$$

and

$$F_\phi^{-1}(x) = \frac{1}{\phi}\log\left[\frac{\sin(\phi\pi x)}{\sin\{\phi\pi(1-x)\}}\right] \qquad (0 < x < 1). \tag{2.5}$$

To generate cluster random effect R_i with bridge distribution, we let $R_i = F_\phi^{-1}(\Phi(Z_i))$ using (2.5), where Φ is the cdf of univariate standard normal distribution and Z_i's are independent and identically distributed with standard normal $N(0,1)$. Under the assumption that R_i has bridge distribution, the marginal probability of success for Y_{1ij} integrated over R_i still has logit link function with regression parameters proportional to the parameters from conditional model (2.1):

$$\text{logit}(p_{1ij}) = X_{ij}^T(\phi\boldsymbol{\alpha}),$$

where population-averaged parameters are attenuated by cluster heterogeneity characterized by attenuation parameter ϕ, which is assumed to be the same for all clusters. In this way, marginal success probability of Y_{1ij} in (2.2) is

$$p_{1ij} = \frac{\exp\{X_{ij}^T(\phi\boldsymbol{\alpha})\}}{1 + \exp\{X_{ij}^T(\phi\boldsymbol{\alpha})\}}. \tag{2.6}$$

The intra-cluster correlation between the binary responses $(Y_{1ij}, Y_{1ij'})$ from two different subjects within the same cluster is $1-\phi$. Small ϕ indicates three important facts: high correlation of the binary responses within a cluster; high heterogeneity among clusters related to the binary response; and a corresponding high degree of attenuation of $\boldsymbol{\alpha}$ in the population due to cluster heterogeneity.

However, the bridge distribution is not closed under linear combination. Hence, when the random intercept logistic regression model is extended to multivariate random effect models, link function for the marginal model does not preserve logit form even each component of the random effects vector has the bridge distribution.

The conditional expectation of Y_{2ij} given R_i after integrated out Y_{1ij} is

$$\begin{aligned} E[Y_{2ij}|R_i, X_{ij}] &= EE[Y_{2ij}|R_i, Y_{1ij}, X_{ij}] \\ &= X_{ij}^T\boldsymbol{\beta} + \gamma(E[Y_{1ij}|R_i, X_{ij}] - p_{1ij}) + R_i \\ &= X_{ij}^T\boldsymbol{\beta} + \gamma(p_{ij} - p_{1ij}) + R_i, \end{aligned}$$

and the marginal expectation of the continuous response Y_{2ij} is

$$E[Y_{2ij}|X_{ij}] = EE[Y_{2ij}|R_i, X_{ij}] = X_{ij}^T\boldsymbol{\beta}$$

since R_i has mean 0. The marginal regression function of Y_{2ij} is again a linear form of X_{ij}.

Hence, both the conditional (given the cluster-specific random effect) and marginal models for the binary and continuous responses of Model 1 have physical interpretations.

2.3 Subject-specific Random Effect Model (Model 2)

As an alternative to Model 1, we propose another single random effect model for the binary and continuous responses, which uses subject-specific random effect instead of cluster-specific random effect in Model 1. In **Model 2**, we replace R_i by Z_{ij} and its transformation R_{ij} using:

$$\text{logit}(p_{ij}) = X_{ij}^T\boldsymbol{\alpha} + R_{ij} \tag{2.7}$$

$$Y_{2ij} = X_{ij}^T\boldsymbol{\beta} + \lambda Z_{ij} + \epsilon_{ij}\ . \tag{2.8}$$

where subject-specific random effect Z_{ij} is assumed to have standard normal distribution, $R_{ij} = F_\phi^{-1}(\Phi(Z_{ij}))$ using (2.5) is again bridge distributed with parameter ϕ, and ϵ_{ij} has

normal distribution with mean 0 and variance σ^2. Within-cluster associations among n_i subjects from cluster i are modeled via the multivariate distribution of $(Z_{i1}, ..., Z_{i,n_i})' \sim MVN(\boldsymbol{\mu}, \Sigma_{\rho_1})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix},$$

and Σ_{ρ_1} is the covariance matrix with variance 1 and common positive correlation ρ_1:

$$\Sigma_{\rho_1} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_1 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_1 & \cdots & \rho_1 & 1 \end{pmatrix}.$$

The standard normal distribution of Z_{ij} and bridge distribution of R_{ij} ensures the marginal models of the two responses satisfy

$$\text{logit}(E[Y_{1ij} = 1|X_{ij}]) = X_{ij}^T(\phi\boldsymbol{\alpha})$$

and

$$E[Y_{2ij}|X_{ij}] = X_{ij}^T\boldsymbol{\beta}.$$

Association of responses Y_{1ij} and $Y_{1ij'}$ and of Y_{2ij} and $Y_{2ij'}$ within the same cluster is induced by correlation ρ_1 between Z_{ij} and $Z_{ij'}$. Association of Y_{1ij} and Y_{2ij} within the same subject is induced by shared $Z_{ij}(R_{ij})$ with coefficient λ, and association between Y_{1ij} and $Y_{2ij'}$ from different subjects within the same cluster is induced by association between R_{ij} (transformation of Z_{ij}) and $Z_{ij'}$ with coefficient parameter λ. Again, this model insures that association between the mixed responses from the same subject is greater than the association between the mixed responses from different subjects within the same cluster. Advantage of Model 2 is that Y_{2ij} is no longer conditioned on the binary response Y_{1ij} in (2.8) and thus marginal distribution of Y_{2ij} still preserves normal density.

2.4 Correlated Bivariate Random Effects Model (Model 3)

A bivariate correlated random effects model is proposed in this section for modeling of clustered data with mixed responses. We use R_{1i} and R_{2i} to denote separate but correlated

cluster-specific random effects for Y_{1ij} and Y_{2ij}, respectively. **Model 3** has the following form:

$$\text{logit}(p_{ij}) = X_{ij}^T\boldsymbol{\alpha} + R_{1i} \tag{2.9}$$

$$Y_{2ij} = X_{ij}^T\boldsymbol{\beta} + \gamma(Y_{1ij} - p_{1ij}) + R_{2i} + \epsilon_{ij}\ , \tag{2.10}$$

where p_{1ij} is the marginal probability of $Y_{1ij} = 1$ defined in (2.6), R_{1i} is assumed to have bridge distribution, and R_{2i} has normal distribution with mean 0 and variance σ_r^2. The marginal regression function of Y_{1ij} is again $\text{logit}(E[Y_{1ij} = 1|X_{ij}]) = X_{ij}^T(\phi\boldsymbol{\alpha})$ (after integrating out R_{1i}), and the marginal mean for Y_{2ij} is again $E[Y_{2ij}|X_{ij}] = X_{ij}^T\boldsymbol{\beta}$ (after integrating out R_{2i} and Y_{1ij}).

The association between the binary outcome Y_{1ij} and the continuous outcome $Y_{2ij'}$ from different subjects within the same cluster is induced by the association between cluster random effects R_{1i} and R_{2i}. Association between the two responses Y_{1ij} and Y_{2ij} within the same subject is induced by association between (R_{1i}, R_{2i}) as well as the dependence of the continuous response Y_{2ij} on the binary response Y_{1ij} via γ. The bivariate density of (R_{1i}, R_{2i}) is constructed through a copula model illustrated as follows.

2.4.1 Copula Model

In statistics, copulas [15] are functions that join multivariate distribution functions to their one-dimensional marginal distribution functions which are uniform, from which dependence between marginal variables can be represented. Sklar's theorem [16] forms the fundamental of most applications of the copula, which states that, if J is a joint distribution function with marginal distribution functions $F_1, ..., F_d$, then there exists a copula C such that the copula binds the margins to the given joint distribution J:

$$J(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)).$$

Moreover, the copula function C is unique if the margins F_i $(1 \leq i \leq d)$ are continuous.

There are many families of copulas depending on the forms of dependence of marginal variables that are represented. Typical copula families includes Gaussian copula, Archimedean copulas, and Periodic copula. We construct the bivariate density of random effects (R_{1i}, R_{2i}) using Gaussian Copula. Let Z_1 and Z_2 be distributed as standard bivariate normal with

correlation ρ, then the Gaussian copula function is

$$C_\rho(U, V) = \Phi_\rho(\Phi^{-1}(U), \Phi^{-1}(V)),$$

where Φ is the cdf of standard normal, $U(Z_1)$ and $V(Z_2)$ are cdfs with range in $(0, 1)$, and

$$\Phi_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right\}$$

is the probability density function for standard bivariate normal with correlation ρ.

By Gaussian Copula, let

$$(R_{1i}, R_{2i}) = (F_\phi^{-1}(\Phi(Z_{1i})), \sigma_r Z_{2i}) \tag{2.11}$$

where $(Z_{1i}, Z_{2i})' \sim MVN(\mathbf{0}, \Sigma_{\rho_2})$ with

$$\Sigma_{\rho_2} = \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix}.$$

R_{1i} and R_{2i} constructed above are correlated through correlation parameter ρ_2 and have the desired marginal distributions (bridge density and normal density, respectively). Since both the bridge cumulative and inverse cumulative distributions have a closed-form expression, generations of R_{1i} and R_{2i} can easily be implemented in standard statistical software and simulations are easy to perform.

2.4.2 Associations Between the Mixed Responses

From (2.10), we get the covariance between the two responses Y_{1ij} and Y_{2ij} from the same subject is

$$\text{Cov}[Y_{1ij}, Y_{2ij}] = \gamma \text{Var}(Y_{1ij}) + \text{Cov}[Y_{1ij}, R_{2i}] \ , \tag{2.12}$$

and the covariance of the two responses from different subjects j and j' within the same cluster i is

$$\text{Cov}[Y_{1ij}, Y_{2ij'}] = \gamma \text{Cov}[Y_{1ij}, Y_{1ij'}] + \text{Cov}[Y_{1ij}, R_{2i}] \tag{2.13}$$

with $\text{Var}(Y_{1ij}) = p_{1ij}(1 - p_{1ij})$.

Our model ensures the anticipated result that

$$|\text{Cov}[Y_{1ij}, Y_{2ij}]| > |\text{Cov}[Y_{1ij}, Y_{2ij'}]|$$

since

$$\text{Var}(Y_{1ij} - Y_{1ij'}) = 2\text{Var}(Y_{1ij}) - 2\text{Cov}[Y_{1ij}, Y_{1ij'}] > 0,$$

from which we get the first term in (2.12) is greater than the first term in (2.13).

For subjects (i, j) and (i, j') with covariate vectors X_{ij} and $X_{ij'}$, we have

$$\text{Cov}[Y_{1ij}, Y_{1ij'}] = E_{R_1}\left[\frac{\exp(X_{ij}^T\boldsymbol{\alpha} + X_{ij'}^T\boldsymbol{\alpha} + 2R_{1i})}{\{1+\exp(X_{ij}^T\boldsymbol{\alpha} + R_{1i})\}\{1+\exp(X_{ij'}^T\boldsymbol{\alpha} + R_{1i})\}}\right] - p_{1ij}p_{1ij'}, \quad (2.14)$$

and

$$\text{Cov}[Y_{1ij}, R_{2i}] = E_{R_1R_2}\left[\frac{\exp(X_{ij}^T\boldsymbol{\alpha} + R_{1i})}{1+\exp(X_{ij}^T\boldsymbol{\alpha} + R_{1i})}R_{2i}\right], \quad (2.15)$$

where E_{R_1} and $E_{R_1R_2}$ are expectations with respect to the marginal density of R_{1i} and with respect to the joint density of (R_{1i}, R_{2i}). Hence, covariance between Y_{1ij} and Y_{2ij} could be computed numerically using (2.12)-(2.15) via simulation study.

Traditionally, association of bivariate data is evaluated by means of correlation coefficient defined as

$$\rho(Y_1, Y_2) = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{\text{Var}(Y_1)\text{Var}(Y_2)}}.$$

A simulation study is conducted to investigate the correlations between random effects R_{1i} and R_{2i} for different values of correlation ρ_2 between latent variables Z_{1i} and Z_{2i}. R_{1i} and R_{2i} are generated using (2.11). We observe that the correlation between the two random intercepts R_{1i} and R_{2i} are approximately the same as the correlation ρ_2 between Z_{1i} and Z_{2i}, as show in Figure 2.1.

The correlation coefficient is a good measure for linear dependence in bivariate normal distribution. However, for general bivariate data which is not bivariate normally distributed, the correlation coefficient usually underestimates the dependence if the association is nonlinear. To understand the associations within (R_{1i}, R_{2i}) and the association within $(Y_{1ij}, Y_{2ij'})$ induced via the correlation ρ_2 of (Z_{1i}, Z_{2i}), we compute Kendall's τ coefficient of (R_{1i}, R_{2i}) and Kendall's τ coefficient of $(Y_{1ij}, Y_{2ij'})$ for different values of ρ_2.

Kendall's τ [17] is a nonparametric rank-based statistic used to measure the degree of correspondence between two rankings, which does not require the normality assumption. Meanwhile, Kendall's τ is unchanged by both linear and nonlinear increasing transformations. However, a drawback of Kendall's τ is that two pairs are needed to interpret τ, while the correlation coefficient requires only one pair.

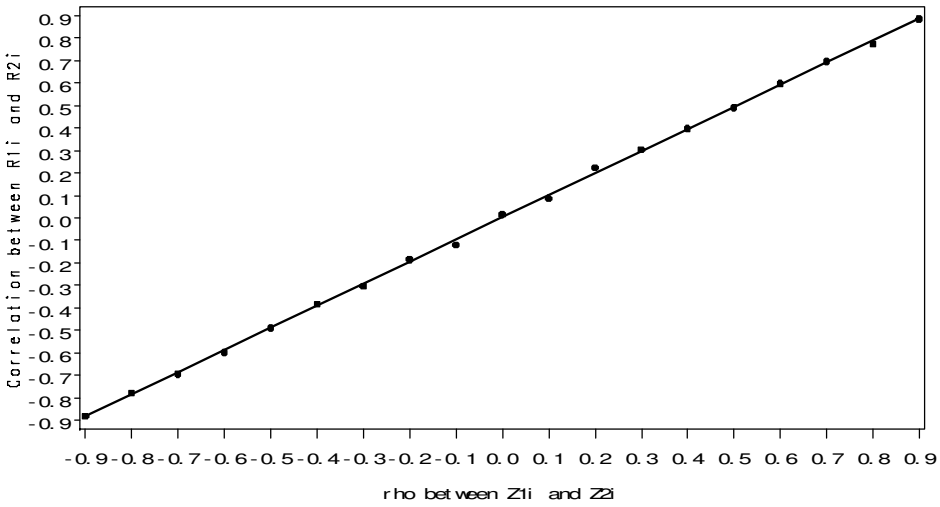

Figure 2.1: Correlation Between R_{1i} and R_{2i} for Different ρ_2

For bivariate data W_{ij}, $i = 1, ..., n$, $j = 1, 2$, which are assumed to be independent and identically distributed for the different values of i, Kendall's τ is defined as

$$\tau = E[\text{sign}\{(W_{11} - W_{21})(W_{12} - W_{22})\}], \tag{2.16}$$

where function sign$(w) = 1$ if $w > 0$, 0 if $w = 0$, and -1 if $w < 0$. For continuous distributions of data, $\tau = 2p - 1$, where

$$p = \text{Prob}[(W_{11} - W_{21})(W_{12} - W_{22}) > 0],$$

which is the probability that the orders of observations from coordinates 1 and 2 are the same.

Kendall's τ for bivariate normal random variables (Z_{1i}, Z_{2i}) with correlation ρ_2 is

$$\tau = 2 \arcsin(\rho_2)/\pi$$

Table 2.1: Parameter Values for Simulating $(Y_{1ij}, Y_{2ij'})$

α_0	α_1	β_0	β_1	γ	ϕ	σ_r	σ
-3.986	1.528	0.945	- 0.076	0	0.735	0.080	0.074

with $-1 \leq \tau \leq 1$. It is also the Kendall's τ for (R_{1i}, R_{2i}) since transformations $R_{1i} = F_b^{-1}(\Phi(Z_{1i}))$ and $R_{2i} = \sigma_1 Z_{2i}$ are both monotone increasing, which Kendall's τ is invariant to.

By (2.16), for bivariate data $(Y_{1ij}, Y_{2ij'})$ which are pairs of the binary outcome and continuous outcome from different subjects within the same cluster, Kendall's τ is defined as

$$E\left[\text{sign}(Y_{1ij} - Y_{1i^*j^*})(Y_{2ij'} - Y_{2i^*j^{**}})\right].$$

We compute Kendall's τ values for simulated data $(Y_{1ij}, Y_{2ij'})$ with different values of ρ_2 via Monte Carlo methods using (2.5), (2.9), (2.10), and (2.11). Table 2.1 shows the values of parameters we used to generated the bivariate data.

Figure 2.2 and Figure 2.3 give the entire range of Kendall's τ of (R_{1i}, R_{2i}) as well as $(Y_{1ij}, Y_{2ij'})$ can be achieved for different values of ρ_2 ranging from $(-0.9, 0.9)$. We observe for every value of ρ_2, there is a correspondent τ value which could be measured as the association of the observed data, and vice versa. This implies the association of mixed responses induced by the cluster-specific random effects could be characterized by ρ_2.

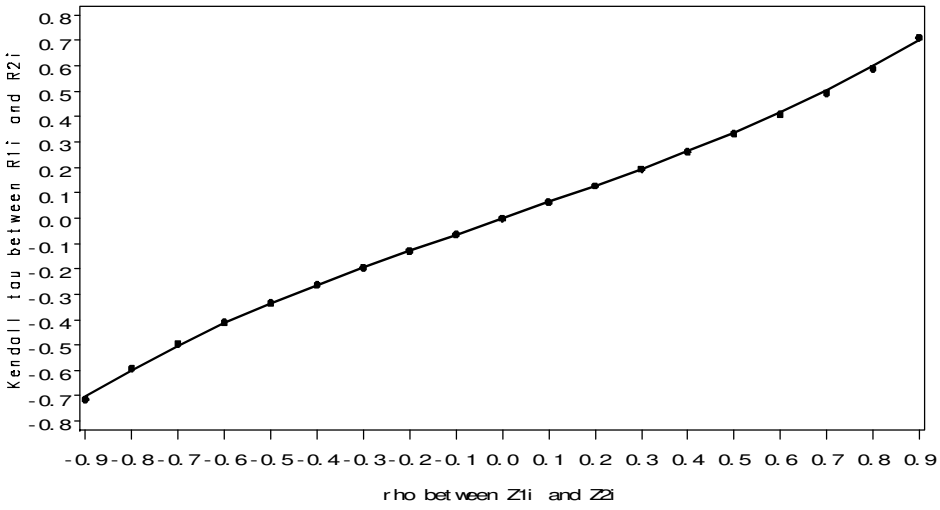

Figure 2.2: Kendall's τ Between R_{1i} and R_{2i} for Different ρ_2

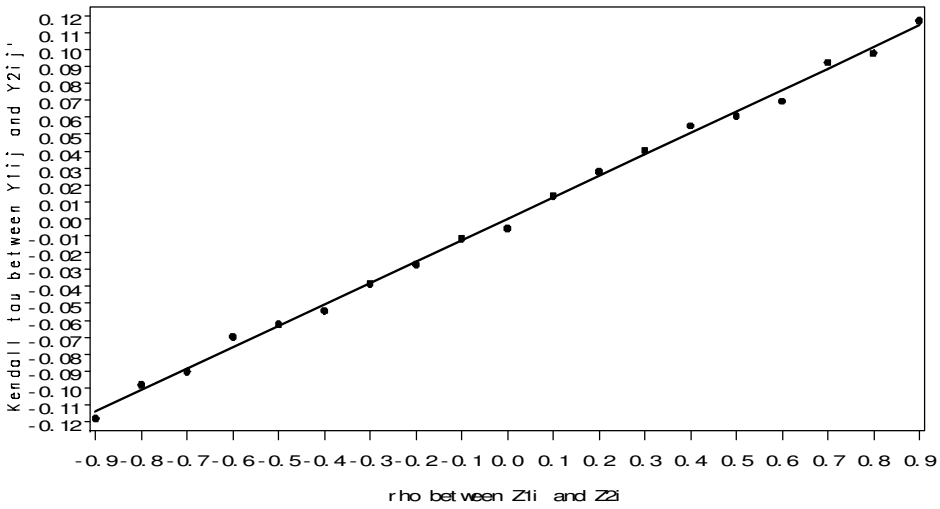

Figure 2.3: Kendall's τ Between Y_{1ij} and $Y_{2ij'}$ for Different ρ_2

CHAPTER 3

PARAMETRIC BAYESIAN ANALYSIS

Two principal frameworks for data analysis are: frequentist approach and Bayesian approach. The frequentist method makes statistical inference based on maximum likelihood estimates (MLEs) and hypothesis tests using p-values; the Bayesian method uses prior for parameters and likelihood for posterior distribution computation, from which inferences are obtained [18]. The major advantages of Bayesian over the frequentist method are: efficiency and accuracy of estimation could be gained by incorporating available information in prior specifications; credible sets on parameters obtained from Bayesian method is more consistent with the natural interpretation of confidence interval compared to frequentist method; complex random effects models are easier to fit using recent sampling methods than classical methods [19].

In the Bayesian method, let $f(\boldsymbol{y}|\boldsymbol{\theta})$ be the distribution for observed data $\boldsymbol{y} = (y_1, ..., y_n)$ conditioned on unknown parameters $\boldsymbol{\theta}$ with prior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ are called hyperparameters. By Bayes' Theorem, the posterior distribution of parameters $\boldsymbol{\theta}$ is [18]

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\xi}) = \frac{p(\boldsymbol{y}, \boldsymbol{\theta}|\boldsymbol{\xi})}{p(\boldsymbol{y}|\boldsymbol{\xi})} = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\xi})}{\int f(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}|\boldsymbol{\xi})d\boldsymbol{u}} \propto f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\xi})$$

Bayesian data analysis requires complex numerical integrations, which could be evaluated by the sampling-based estimation methods developed recently, e.g. Marcov Chain Monte Carlo (MCMC) sampling method. In this Chapter, we first review MCMC techniques as well as some major sampling methods which will be used later for data analysis. We also describe fully parametric Bayesian approach to analyze the single random effect models and correlated random effect models (Model 1, Model 2 and Model 3).

3.1 MCMC and Sampling Methods

3.1.1 Basic Concepts of MCMC

We give the basic concepts and idea behind MCMC method in this section. Most of the contents in this section are taken from Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference [20] and Computational Methods in Statistics [21].

A **stochastic process** is a set $\{\theta^{(t)} : t \in T\}$ with state space S and index set T.

A **Markov Chain** is a stochastic process where given the present state, past and future states are independent:

$$f_{\theta^{(n+1)}|\theta^{(n)},...,\theta^{(0)}}(\vartheta^{(n+1)}|\vartheta^{(n)},...,\vartheta^{(0)}) = f_{\theta^{(n+1)}|\theta^{(n)},...,\theta^{(0)}}(\vartheta^{(n+1)}|\vartheta^{(n)}), \tag{3.1}$$

where $\vartheta^{(0)},...,\vartheta^{(n+1)} \in S$. The conditional probability density in (3.1) depends on $\vartheta^{(n)}$, $\vartheta^{(n+1)}$ and n.

A Markov process is called **homogeneous** if the conditional probability density dose not dependent on n:

$$f_{\theta^{(n+1)}|\theta^{(n)}}(\vartheta^{(n+1)}|\vartheta^{(n)}) = f_{\theta^{(1)}|\theta^{(0)}}(\vartheta^{(1)}|\vartheta^{(0)}).$$

A fundamental problem of Markov chain simulation is the asymptotic behavior of the chain as $n \to \infty$. A stochastic process is called **stationary** if its n^{th} order joint density function is translation invariant:

$$f_{\theta^{(0)},\theta^{(1)},...,\theta^{(n)}}(\vartheta^{(0)},\vartheta^{(1)},...,\vartheta^{(n)}) = f_{\theta^{(c)},\theta^{(c+1)},...,\theta^{(c+n)}}(\vartheta^{(c)},\vartheta^{(c+1)},...,\vartheta^{(c+n)}).$$

The underlying principle of MCMC methods is to start with a homogeneous Markov process which is not stationary, but converge to a stationary process as $n \to \infty$, of which the marginal density is the desired probability density, then the values sampled from this process become the samples from this density as n increases to some large value.

In the following, we give the conditions for a Markov chain converges to a unique stationary distribution for finite state space, infinite countable state space, and uncountable state space.

Assume the desired probability distribution is on a finite discrete state space $S = \{s_1,...,s_m\}$. Let $\Pi_{ij} = \Pr(\theta^{(n)} = s_j|\theta^{(n-1)} = s_i)$, then the $m \times m$ **transition probability**

matrix is defined as

$$\Pi = \begin{pmatrix} \Pi_{11} & \cdots & \Pi_{1m} \\ \cdots & \cdots & \cdots \\ \Pi_{m1} & \cdots & \Pi_{mm} \end{pmatrix}$$

with each row adds up to 1. The probability of transition from s_i to s_j in n steps is given by $(\Pi^n)_{ij}$.

Peron-Frobenius Theorem states that if $\Pi^n >> 0$ for some non-negative integer n, then there exists a $X >> 0$ such that $X\Pi = X$. This theorem implies that if there exists an n such that an arbitrary state can to to any other state in n steps with positive probability, then the resulting Markov chain has a unique stationary probability vector P, which is obtained by normalizing X into a probability vector $P = X/(\sum_{i=1}^{m} X_i)$.

Two states are said to **communicate** if one state could go to another in a finite number of steps, i.e. s_i and s_j are communicate if there exists a positive integer n such that $(\Pi^n)_{ij} > 0$.

A Markov chain is said to be **irreducible** if all states in S communicate with each other for the corresponding transition matrix Π.

For a state s_i and a given transition matrix Π, the **period** of s_i is defined as

$$d(s_i) = GCD\{n : (\Pi^n)_{ii} > 0\},$$

where GCD denotes the greatest common divisor. If two states communicate, then they have the same periods.

An irreducible Markov chain is called **aperiodic** if its period is one.

An irreducible, aperiodic, homogeneous Markov chain on a finite state space has the property that $\Pi^n >> 0$ for some $n > 0$. According to Peron-Frobenious theorem, for any arbitrary starting condition, such a Markov chain converges to a stationary process and generates samples from P, a unique probability distribution such that $P\Pi = P$, as the time goes to infinity.

For infinite countable state space χ, the Markov chain convergence requires an additional property of recurrence to ensure the chain has a unique stationary probability.

For a state $s_i \in \chi$, f_{ii}^n is defined as the probability that a homogeneous Markov chain revisits s_i for the first time in n steps:

$$f_{ii}^n = \Pr(\theta^{(n)} = s_i, \theta^{(n-1)} \neq s_i, ..., \theta^{(1)} \neq s_i | \theta^{(0)} = s_i).$$

A state $s_i \in \chi$ is called **recurrence** under a Markov chain if and only if $\Sigma_{n=1}^{\infty}\Pi_{ii}^{n} = \infty$.

Let $P_i = \lim_{n\to\infty}(\Pi^n)_{ii}$, if $P_i > 0$ for the state s_i in an aperiodic, recurrent class, then $P_j = \lim_{n\to\infty}(\Pi^n)_{ij}$ for all the states s_i in that class. This class is called **positive recurrent**. A Markov chain is called positive recurrent if it is irreducible, aperiodic, recurrent, and $P_i > 0$ for at least one state in the class.

On a countable state space, an irreducible, aperiodic, positive recurrent Markov chain has a unique stationary probability function.

For a uncountable State Space χ, let $\tau_A = \inf\{n \geq 1 | \theta^{(n)} \in A\}$, the time of first revisit to A. A Markov chain is **ϕ-irreducible** if there exists a non-zero measure ϕ on χ such that $\Pr(\tau_A < \infty | \theta^{(0)} = x) > 0$ for all $x \in A$ and $A \subset \chi$ that satisfy $\phi(A) > 0$.

A $\phi-$irreducible Markov chain with a stationary probability density $f(x)$ is **Harris recurrent** if for all $A \subset \chi$ with $f(A) > 0$ and for all $x \subset \chi$, $\Pr(\tau_A < \infty | \theta^{(0)} = x) = 1$.

On a uncountable state space, a $\phi-$irreducible, aperiodic, Harris recurrent Markov chain has a unique stationary density f.

3.1.2 Review of Major Sampling Methods

In the following, we review the adaptive rejection method developed by Gilks and Wild [22], the slice sampling method proposed by Neal [23], the Metropolis-Hastings algorithm proposed by Metropolis and Hastings [24] [25], the Gibbs sampling algorithm developed by Geman and Geman [26], and the Metropolis-within-Gibbs sampling method suggested by Muller [27].

Adaptive rejection method [22] is a resampling scheme for log-concave densities. Assume that $\pi(\theta)$ is continuous and differentiable and $h(\theta) = \ln \pi(\theta)$ is concave everywhere on domain D. Let $T_m = (\theta^1, ..., \theta^m)$ be the m starting points with $\theta^1 \leq ... \leq \theta^m$, where θ^1 satisfies $h'(\theta^1) > 0$ if D is unbounded on the left and θ^m satisfies $h'(\theta^m) < 0$ if D is unbounded on the right. Let $u_m(\theta)$ be the piece-wise linear upper bound formed from the tangents to $h(\theta)$ at each point in T_m, defined as:

$$u_m(\theta) = h(\theta^i) + (\theta - \theta^i)h'(\theta^i) \tag{3.2}$$

for $\theta \in [z^{i-1}, z^i]$ $(i = 1, ..., m)$ with z^i as the point at which the tangent lines at θ^i and θ^{i+1} intersects for $i = 1, ..., m-1$:

$$z^i = \frac{h(\theta^{i+1}) - h(\theta^i) - \theta^{i+1}h'(\theta^{i+1}) + \theta^i h'(\theta^i)}{h'(\theta^i) - h'(\theta^{i+1})},$$

and z^0 as the lower bound of D ($-\infty$ if D is unbounded on the left). Define

$$s_m(\theta) = \frac{\exp(u_m(\theta))}{\int_D \exp(u_m(\theta'))d\theta'}. \tag{3.3}$$

The piece-wise linear lower bound $l_m(\theta)$ is defined similarly:

$$l_m(\theta) = \frac{(\theta^{i+1} - \theta)h(\theta^i) - (\theta - \theta^i)h(\theta^{i+1})}{\theta^{i+1} - \theta^i} \tag{3.4}$$

for $\theta \in [\theta^i, \theta^{i+1}]$ $(i = 1, ..., m-1)$ and $l_m(\theta) = -\infty$ if $\theta < \theta^1$ or $\theta > \theta^m$.

To sample n points independently from $\pi(\theta)$, the algorithm is set up as follows:

1. Evaluate $u_m(\theta)$, $l_m(\theta)$ and $s_m(\theta)$ at initial points T_m according to (3.2), (3.4) and (3.3);

2. Draw a sample θ^* from $s_m(\theta)$ and a random value u from $U(0, 1)$. If $u \leq \exp(l_m(\theta^*) - u_m(\theta^*))$ which is the upper envelop function, then accept θ^*, otherwise calculate $h(\theta^*)$ and $h'(\theta^*)$. If $u \leq \exp(h(\theta^*) - u_m(\theta^*))$ which is the rejection envelop function, then accept θ^*, otherwise reject θ^*;

3. If $h(\theta^*)$ and $h'(\theta^*)$ were evaluated in step 2, include θ^* in T_m to form T_{m+1}. Reorder elements in T_{m+1} ascendingly. Reconstruct $u_{m+1}(\theta)$, $l_{m+1}(\theta)$ and $s_{m+1}(\theta)$;

4. Repeat step 2 and step 3 until n points are accepted.

Derivative-free adaptive rejection method [28] extends the adaptive rejection method such that it does not require the existence of continuous derivatives. The envelope function is constructed in a different way which does not require the evaluation of derivatives.

If $\pi(\theta)$ is not log-concave but restricted to a finite range, we consider the **slice sampling** [23] method. Suppose function $f(\theta)$ is proportional to density $\pi(\theta)$, the slice sampling algorithm to replace current value $\theta^{(0)}$ by $\theta^{(1)}$ has the following three-step procedure:

1. Generate a random value θ^* uniformly from $(0, f(\theta^{(0)}))$. A horizontal slice is defined by $S = \{\theta : \theta^* < f(\theta)\}$. $\theta^{(0)}$ is always within S by definition;

2. Search an interval I around $\theta^{(0)}$ that contains all or much of the slice as follows: randomly positioned an interval with width w around $\theta^{(0)}$, then expand the interval in steps of size w until both ends of the interval are outside the slice;

3. Sample the new point $\theta^{(1)}$ to replace $\theta^{(0)}$ by picking a value uniformly from the interval I until a point inside the slice is obtained. Points picked outside the slice are used to shrink the interval I.

When distribution $\pi(\theta)$ is not log-concave and unbounded, **Metropolis-Hastings algorithms** [24] [25] can be used. Assume a transition kernel $q(\theta, \vartheta)$ satisfies

1. $q(\theta, \cdot)$ could be sample from for all θ;
2. The support of q contains the support of π;
3. The functional form of $q(\theta, \vartheta)$ is known or $q(\theta, \vartheta)$ is symmetric in θ and ϑ.

Define the acceptance probability as

$$\alpha(\theta, \vartheta) = \min\left\{1, \frac{\pi(\vartheta)q(\vartheta, \theta)}{\pi(\theta)q(\theta, \vartheta)}\right\}. \tag{3.5}$$

then the Metropolis-Hastings algorithm is described as follows:

1. Set the iteration counter $j = 1$ and generate initial values for $\theta^{(0)}$;
2. Generate a new ϑ from the density $q(\theta^{(j-1)}, \cdot)$;
3. Calculate the acceptance probability of $\alpha(\theta^{(j-1)}, \vartheta)$ using (3.5). Update $\theta^{(j-1)}$ state to $\theta^{(j)}$ according to:

$$\theta^{(j)} := \begin{cases} \vartheta & \text{with probability } \alpha(\theta^{(j-1)}, \vartheta), \\ \theta^{(j-1)} & \text{with probability } 1 - \alpha(\theta^{(j-1)}, \vartheta); \end{cases}$$

4. Set $j = j + 1$ and repeat step 2 to 3 until convergence is reached.

Step 3 is implemented using a random quantity u generated from a uniform distribution on $(0, 1)$. If $u \leq \alpha$ the move is accepted, otherwise $\theta^{(j)}$ have the same value as $\theta^{(j-1)}$.

Let $\pi(\boldsymbol{\theta})$ be the multivariate density of parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)$, where θ_i $(i = 1, ..., d)$ can be a scalar, a vector or a matrix. Assume the full conditional distributions $\pi(\theta_i|\theta_{-i})$ is known and can be sampled from. The fundamental principle of **Gibbs sampling** [26] is based on successive sampling from the full conditional distribution of each θ_i given all the others. The Gibbs sampling algorithm is set up as follows:

1. Set the iteration counter $j = 1$ and generate initial values for $\boldsymbol{\theta^{(0)}} = (\theta_1^{(0)}, ..., \theta_d^{(0)})$;
2. Generate a new $\boldsymbol{\theta^{(j)}} = (\theta_1^{(j)}, ..., \theta_d^{(j)})$ from $\boldsymbol{\theta^{(j-1)}}$ by successively sample values

$$\theta_1^{(j)} \sim \pi(\theta_1|\theta_2^{(j-1)}, ..., \theta_d^{(j-1)}),$$

$$\theta_2^{(j)} \sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, ..., \theta_d^{(j-1)}),$$

...

$$\theta_d^{(j)} \sim \pi(\theta_d|\theta_1^{(j)}, \theta_2^{(j)}, ..., \theta_d^{(j)});$$

3. Set $j = j + 1$ and repeat step 2 until convergence is reached.

Values of $\boldsymbol{\theta}^{(j)}$ are considered samples from π when the Markov chain converges after a long run time.

The **Metropolis-within-Gibbs sampling** method [27] combines Metropolis-Hastings method and Gibbs sampling method to solve the problem that some components of parameter $\boldsymbol{\theta}$ cannot be directly drawn from their full conditional distribution using Gibbs sampling. The basic idea behinds Metropolis-within-Gibbs sampling is that a Metropolis step is used on some or all components which are sampled from transition kernel q with acceptance probability by (3.5), instead of exact full conditional distribution, inside a Gibbs sampling circle. Each iteration of Gibbs sampling contains d Metropolis-Hastings samplings for d components of $\boldsymbol{\theta}$. The process continues until convergence.

3.2 Parametric Bayesian Method for Model 1

We specify priors and derive likelihood and posterior distribution for Model 1, Model 2 and Model 3 in the following.

3.2.1 Prior Specification

Parameters in Model 1 include $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \phi, \sigma^2)$ with $\boldsymbol{\theta_1} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \lambda)$. Attenuation parameter ϕ of bridge distribution has marginal noninformative prior uniform$(0, 1)$. Precision parameter $1/\sigma^2$ is assigned with gamma marginal prior distribution. Assume all the parameters are independent, then the joint prior density is

$$\pi(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta_1})\pi_2(\phi)\pi_3(\sigma^2), \tag{3.6}$$

where $\pi_1(\boldsymbol{\theta_1})$ has the density of $MVN(\boldsymbol{\mu}, \Sigma)$ with corresponding means and diagonal covariance matrix Σ_1, $\pi_2(\phi)$ is $U(0, 1)$, and $\pi_3(\sigma^2)$ is inverse gamma. Detailed specifications of priors for the data example will be discussed later.

3.2.2 Likelihood and Posterior

The binary response Y_{1ij} conditioned on random effect R_i and continuous response Y_{2ij} given R_i and Y_{1ij} are considered independent. Also Y_{1ij}'s are conditional independent given R_i, and Y_{2ij}'s are independent conditioned on R_i and Y_{1ij}. The likelihood contribution from jth subject from ith cluster is

$$\begin{aligned} &L(\boldsymbol{\theta_1}, \sigma^2, R_i | Y_{1ij}, Y_{2ij}) = f_{Y_{1ij}|R_i}(y_{1ij}|r_i) \times f_{Y_{2ij}|R_i, Y_{1ij}}(y_{2ij}|r_i, y_{1ij}) \\ =&\frac{\exp\{(x_{ij}^T\boldsymbol{\alpha} + r_i)y_{1ij}\}}{\{1+\exp(x_{ij}^T\boldsymbol{\alpha} + r_i)\}} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\left\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - \lambda r_i\right\}^2}{2\sigma^2}\right]. \end{aligned} \tag{3.7}$$

Thus we have the full likelihood of $\boldsymbol{\theta}$ based on data $(\boldsymbol{Y_1}, \boldsymbol{Y_2})$ is

$$L(\boldsymbol{\theta}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) = \prod_{i=1}^{N} \int_{R_i} \left\{ \prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \sigma^2, R_i | Y_{1ij}, Y_{2ij}) \right\} f_{R_i}(r_i|\phi) dR_i,$$

where $f_{R_i}(r_i|\phi)$ is the bridge density given in (2.4).

Let $\boldsymbol{R} = (R_1, ..., R_N)$, the joint posterior distribution for parameters $(\boldsymbol{\theta}, \boldsymbol{R})$ is

$$P(\boldsymbol{\theta}, \boldsymbol{R}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \ \propto \left[\prod_{i=1}^{N} \left\{ \prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \sigma^2, R_i | Y_{1ij}, Y_{2ij}) \right\} f_{R_i}(r_i|\phi) \right] \pi(\boldsymbol{\theta}),$$

where $\pi(\boldsymbol{\theta})$ is given as (3.6) and $L(\boldsymbol{\theta_1}, \sigma^2, R_i | Y_{1ij}, Y_{2ij})$ is given by (3.7). The posterior inference can be addressed via MCMC samples from $P(\boldsymbol{\theta}, \boldsymbol{R}|\boldsymbol{Y_1}, \boldsymbol{Y_2})$.

3.3 Parametric Bayesian Method for Model 2

3.3.1 Prior Specification

For Model 2, we have $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \sigma^2)$ with $\boldsymbol{\theta_1} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$ and $\boldsymbol{\theta_2} = (\phi, \rho_1)$. The marginal prior for $\boldsymbol{\theta_1}$ is again a MVN. Attenuation parameter ϕ of bridge distribution and positive correlation parameter ρ_1 has marginal noninformative prior uniform$(0, 1)$. Precision parameter $1/\sigma^2$ is assigned with gamma marginal prior distribution. Joint prior density is then assumed to be

$$\pi(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta_1})\pi_2(\phi)\pi_3(\rho_1)\pi_4(\sigma^2), \tag{3.8}$$

where $\pi_1(\boldsymbol{\theta_1})$ has the density of $MVN(\boldsymbol{\mu}, \Sigma)$ with corresponding means and diagonal covariance matrix Σ_2, $\pi_2(\phi)$ is $U(0, 1)$, $\pi_3(\rho_1)$ is $U(0, 1)$, and $\pi_4(\sigma^2)$ is inverse gamma.

3.3.2 Likelihood and Posterior

The binary response Y_{1ij} and continuous response Y_{2ij} conditioned on subject-specific random effects R_{ij} and Z_{ij} are considered independent. Also Y_{1ij}'s are conditional independent given R_{ij}, and Y_{2ij}'s are independent conditioned on Z_{ij}. The likelihood contribution from jth subject from ith cluster is

$$\begin{aligned} L(\boldsymbol{\theta_1}, \phi, \sigma^2, Z_{ij}|Y_{1ij}, Y_{2ij}) &= f_{Y_{1ij}|R_{ij}}(y_{1ij}|r_{ij}) \times f_{Y_{2ij}|Z_{ij}}(y_{2ij}|z_{ij}) \\ &= \frac{\exp\{(x_{ij}^T\boldsymbol{\alpha} + r_{ij})y_{1ij}\}}{\{1+\exp(x_{ij}^T\boldsymbol{\alpha} + r_{ij})\}} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \lambda z_{ij})^2}{2\sigma^2}\right\} \end{aligned} \tag{3.9}$$

with $R_{ij} = F_\phi^{-1}(\Phi(Z_{ij}))$.

Let $\boldsymbol{Z_i} = (Z_{i,1}, ..., Z_{i,n_i})$ $(i = 1, ..., N)$, then the full likelihood of $\boldsymbol{\theta}$ based on data $(\boldsymbol{Y_1}, \boldsymbol{Y_2})$ is

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) = &\prod_{i=1}^{N} \int_{Z_{i,n_i}} \cdots \int_{Z_{i1}} \left\{\prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \phi, \sigma^2, Z_{ij}|Y_{1ij}, Y_{2ij})\right\} f_{\boldsymbol{Z_i}}(z_{i1}, \cdots, z_{i,n_i}|\rho_1) \\ &dZ_{i1} \cdots dZ_{i,n_i}, \end{aligned}$$

where $f_{\boldsymbol{Z_i}}(z_{i1}, \cdots, z_{i,n_i}|\rho_1)$ is the multivariate standard normal density with mean 0, variance 1 and correlation ρ_1.

Let $\boldsymbol{Z} = (\boldsymbol{Z_1}, ..., \boldsymbol{Z_N})$, the joint posterior distribution for parameters $(\boldsymbol{\theta}, \boldsymbol{Z})$ is

$$P(\boldsymbol{\theta}, \boldsymbol{Z}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \propto \left[\prod_{i=1}^{N} \left\{\prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \phi, \sigma^2, Z_{ij}|Y_{1ij}, Y_{2ij})\right\} f_{\boldsymbol{Z_i}}(z_{i1}, \cdots z_{i,n_i}|\rho_1)\right] \pi(\boldsymbol{\theta}),$$

where $\pi(\boldsymbol{\theta})$ is given as (3.8) and $L(\boldsymbol{\theta_1}, \phi, \sigma^2, Z_{ij}|Y_{1ij}, Y_{2ij})$ is given by (3.9). Again, the posterior inference can be addressed via MCMC samples from $P(\boldsymbol{\theta}, \boldsymbol{Z}|\boldsymbol{Y_1}, \boldsymbol{Y_2})$.

3.4 Parametric Bayesian Method for Model 3

3.4.1 Prior Specification

Parameters in models (2.9) and (2.10) include $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \sigma^2)$ with $\boldsymbol{\theta_1} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ and $\boldsymbol{\theta_2} = (\phi, \rho_2, \sigma_r^2)$, where $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ has $MVN(\boldsymbol{\mu}, \Sigma)$ prior, ϕ and the correlation ρ_2 of (Z_{1i}, Z_{2i}) have vague priors uniform$(0, 1)$ and uniform$(-1, 1)$, and both precision parameters $1/\sigma_r^2$ and $1/\sigma^2$ have gamma prior. Joint prior density is then given by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta_1})\pi(\boldsymbol{\theta_2})\pi(\sigma^2) = \pi_1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)\pi_2(\phi)\pi_3(\rho_2)\pi_4(\sigma_r^2)\pi_5(\sigma^2). \tag{3.10}$$

3.4.2 Likelihood and Posterior

The likelihood contribution from jth subject from ith cluster is

$$\begin{aligned} &L(\boldsymbol{\theta_1}, \sigma^2, R_{1i}, R_{2i}|Y_{1ij}, Y_{2ij}) = f_{Y_{1ij}|R_{1i}}(y_{1ij}|r_{1i}) \times f_{Y_{2ij}|R_{2i},Y_{1ij}}(y_{2ij}|r_{2i}, y_{1ij}) \\ =&\frac{\exp\{(x_{ij}^T\boldsymbol{\alpha} + r_{1i})y_{1ij}\}}{\{1+\exp(x_{ij}^T\boldsymbol{\alpha} + r_{1i})\}} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\left\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\right\}^2}{2\sigma^2}\right]. \end{aligned} \tag{3.11}$$

And the full likelihood of $\boldsymbol{\theta}$ based on data $(\boldsymbol{Y_1}, \boldsymbol{Y_2})$ is

$$L(\boldsymbol{\theta}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) = \prod_{i=1}^{N} \int_{R_{1i}\times R_{2i}} \left\{\prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \sigma^2, R_{1i}, R_{2i}|Y_{1ij}, Y_{2ij})\right\} f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2}) dR_{1i} dR_{2i}, \tag{3.12}$$

where $f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2})$ is the joint density of (R_{1i}, R_{2i}).

For $R_{1i} = F_\phi^{-1}(\Phi(Z_{1i}))$ and $R_{2i} = \sigma_r Z_{2i}$, we have

$$Z_{1i} = h_1(R_{1i}) = \Phi^{-1}(F_\phi(R_{1i}))$$

and

$$Z_{2i} = h_2(R_{2i}) = R_{2i}/\sigma_r.$$

Assume $\boldsymbol{Z_i} = (Z_{1i}, Z_{2i})$ has bivariate standard normal distribution with correlation ρ_2, the joint probability density function of (R_{1i}, R_{2i}) is obtained using

$$f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}) = f_{\boldsymbol{Z_i}}(h_1(r_{1i}), h_2(r_{2i}))|\boldsymbol{J}|$$

where $f_{\boldsymbol{Z_i}}(\cdot,\cdot)$ is the bivariate normal density with mean 0, variance 1 and correlation ρ_2. Hence $f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2})$ is given as

$$f_{\boldsymbol{Z_i}}(\Phi^{-1}(F_\phi(r_{1i})), \frac{r_{2i}}{\sigma_r}) \left\{\frac{f_\phi(r_{1i})}{f_N(\Phi^{-1}(F_\phi(r_{1i})))\sigma_r}\right\}, \tag{3.13}$$

with $f_N(\cdot)$ being the univariate standard normal density.

Let $\boldsymbol{R_1} = (R_{11}, ..., R_{1N})$ and $\boldsymbol{R_2} = (R_{21}, ..., R_{2N})$, the joint posterior distribution for parameters is $(\boldsymbol{\theta}, \boldsymbol{R_1}, \boldsymbol{R_2})$ and posterior inference can be addressed via MCMC samples from

$$P(\boldsymbol{\theta}, \boldsymbol{R_1}, \boldsymbol{R_2}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \propto \left[\prod_{i=1}^{N} \left\{\prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \sigma^2, R_{1i}, R_{2i}|Y_{1ij}, Y_{2ij})\right\} f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2})\right] \pi(\boldsymbol{\theta})$$

using (3.10), (3.11) and (3.13).

3.5 Model Selection

The proposed three models have their own advantages and shortcomings. For example, the correlated random effects model uses two correlated yet separate cluster random effects for the binary and continuous responses within a cluster. This assumption may contribute to a bigger dimensional model than what is necessary for a study at hand, while the single random effect models are more parsimonious. A good model based on essentially are cluster random effect for both types of responses, if achievable and adequate for an application, will give a model with smaller dimension and easier physical interpretations. Also, the marginal distribution for the continuous response preserves normal density from Model 2, while the marginal functions for the continuous response from Model 1 and Model 3 preserve linear forms but their marginal distributions are not again normal. Because we are considering multiple models for analysis, a careful and appropriate model selection method is very necessary.

Predictive approach is used for Bayesian model determination among the single random effect models and correlated random effects model assessed by conditional predictive ordinate (CPO) statistic. CPO is a cross-validated probability which gives the marginal posterior predictive density of each point given the rest of the observed data. For more discussion of CPO statistic we refer reader to Gelfand, Dey, and Chang [29], and Ibrahim, Chen, and Sinha [30].

Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$ be the $n \times 1$ data vector, y_i be the observed value of the ith response Y_i, and $\boldsymbol{Y_{-i}}$ be the $(n-1) \times 1$ data vector with Y_i excluded. The CPO statistic for unit i is defined as:

$$CPO_i = f(y_i|\boldsymbol{y}_{-i}) = \int f(y_i|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\boldsymbol{y}_{(-i)}) d\boldsymbol{\theta}, \tag{3.14}$$

where $\boldsymbol{\theta}$ is the parameter vector, y_i is the observed data for unit i, and $\boldsymbol{y}_{-i}$ is the observed

data vector $\boldsymbol{y}$ except unit i. Then we have

$$\begin{aligned}
(CPO_i)^{-1} &= \left\{ \int f(y_i|\boldsymbol{\theta}) \frac{f(\boldsymbol{y}_{-i}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{y}_{-i}|\boldsymbol{\theta'})\pi(\boldsymbol{\theta'})d\boldsymbol{\theta'}} d\boldsymbol{\theta} \right\}^{-1} \\
&= \frac{\left\{ \int f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \right\}^{-1}}{\left\{ \int f(\boldsymbol{y_{-i}}|\boldsymbol{\theta'})\pi(\boldsymbol{\theta'})d\boldsymbol{\theta'} \right\}^{-1}} \\
&= \frac{\int f(\boldsymbol{y_{-i}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\boldsymbol{y}|\boldsymbol{\theta'})\pi(\boldsymbol{\theta'})d\boldsymbol{\theta'}} \\
&= \int \frac{1}{f(y_i|\boldsymbol{\theta})} \cdot \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta'})\pi(\boldsymbol{\theta'})d\boldsymbol{\theta'}} d\boldsymbol{\theta} \\
&= \int \{f(y_i|\boldsymbol{\theta})\}^{-1} P(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \\
&= E_{\boldsymbol{\theta}} \left[\{f(y_i|\boldsymbol{\theta})\}^{-1} | \boldsymbol{y} \right].
\end{aligned}$$

Thus the CPO value could be estimated easily using output from a standard MCMC run by

$$CPO_i^{-1} = \frac{1}{k} \sum_{k=1}^{K} \{f(y_i|\boldsymbol{\theta}^{(k)})\}^{-1}$$

for length K large enough. The CPO statistics could be used to identify influential point or outlier which has especially low CPO value.

Since the basic exchangeable unit for our modeled data is a cluster, we compute CPO values for each cluster using (e.g. for Model 3)

$$CPO_{i,M_3} = \left\{ E \left[\frac{1}{\prod_j f_{Y_{1ij}|R_{1i}}(y_{1ij}|r_{1i}) f_{Y_{2ij}|R_{2i},Y_{1ij}}(y_{2ij}|r_{2i}, y_{1ij})} \right] \right\}^{-1}, \tag{3.15}$$

where the expectation is taken with respect to the posterior density of Model 3. To compare the three competing models, we evaluate CPO values under all the models. As an overall summary measure, we use $CPO_{M_k} = \prod_i CPO_{i,M_k}$ $(k = 1, 2, 3)$ as the CPO value for Model k. Data with larger CPO value supports the the corresponding model better than other models.

CHAPTER 4

APPLICATION: EG DATA

We apply the parametric Bayesian approach described in Chapter 3 to the three random effects models proposed in Chapter 2 for real data analysis and compare the models based on CPO statistics.

4.1 Data Description

Our research is motivated by data from the development toxicity study of Ethylene Glycol (EG) conducted through the National Toxicology Program (NTP) [5]. During the period of major organogenesis just after implantation, four different levels of EG, 0, 750, 1500, or 3000 mg/kg/day, were administered to 94 pregnant mice (dams). Totally 1028 fetuses from 94 litters (clusters), with cluster size ranging from 1 to 16, are measured for malformation and weight as the binary and continuous responses. Figure 4.1 shows the hierarchical structure of EG data.

Our aim of the study is to evaluate the effects of EG and within-cluster and within-subject effects on the binary and continuous responses, as well as the associations between binary outcomes in the same cluster, the associations between continuous outcomes in the same cluster, the associations between binary and continuous outcomes from different subjects within a cluster, and the associations between binary and continuous outcomes from the same subject.

Each dose level was administered to about 22-25 litters, with total 226-297 fetuses. From Table 4.1, we clearly observe that the proportion of malformation increases from 0.34% to 57.08% and sample mean of weight decreases from 0.972 to 0.704 as dose level of EG increases from 0 to 3 g/kg/day.

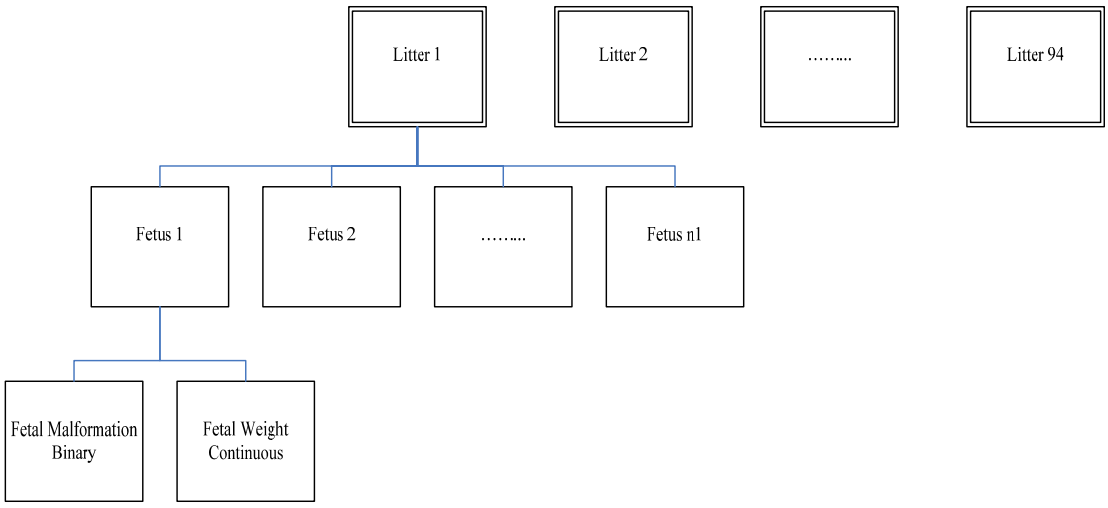

Figure 4.1: Hierarchical Structure of EG Data

4.2 Prior Elicitation

Treating dose X_i $(=X_{ij})$ assigned to the ith pregnant mouse/litter as a covariate, Y_{1ij} as the binary response (fetal malformation) and Y_{2ij} as the continuous response (fetal weight), the three random effects models applied to EG data have the following forms:

Model 1:

$$\text{logit}(E[Y_{1ij}|R_i, X_i]) = \alpha_0 + \alpha_1 X_i + R_i$$
$$Y_{2ij} = \beta_0 + \beta_1 X_i + \gamma(Y_{1ij} - p_{1i}) + \lambda R_i + \epsilon_{ij},$$

where the cluster-specific random effect R_i has bridge distribution with parameter ϕ, $p_{1i} = \exp\{\phi(\alpha_0 + \alpha_1 X_i)\}/[1 + \exp\{\phi(\alpha_0 + \alpha_1 X_i)\}]$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Table 4.1: Data Description for Developmental Toxicity Study of EG

Dose (g/kg)	Live	Litters	Malformations No.(%)	Fetal Weight(g) Mean (S.D.)
0	297	25	1 (0.34)	0.972 (0.098)
0.75	276	24	26 (9.42)	0.877 (0.104)
1.5	229	22	89 (38.86)	0.764 (0.107)
3	226	23	129 (57.08)	0.704 (0.124)

Model 2:

$$\text{logit}(E[Y_{1ij}|R_{ij}, X_i]) = \alpha_0 + \alpha_1 X_i + R_{ij}$$
$$Y_{2ij} = \beta_0 + \beta_1 X_i + \lambda Z_{ij} + \epsilon_{ij},$$

where the subject-specific random effect Z_{ij} has standard normal distribution and its transformation R_{ij} has bridge distribution with parameter ϕ, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Model 3:

$$\text{logit}(E[Y_{1ij}|R_{1i}, X_i]) = \alpha_0 + \alpha_1 X_i + R_{1i}$$
$$Y_{2ij} = \beta_0 + \beta_1 X_i + \gamma(Y_{1ij} - p_{1i}) + R_{2i} + \epsilon_{ij}; \tag{4.1}$$

where the cluster-specific random effects R_{1i} has bridge distribution with parameter ϕ and R_{2i} has normal distribution with mean 0 and variance σ_r^2, $p_{1i} = \exp\{\phi(\alpha_0 + \alpha_1 X_i)\}/[1 + \exp\{\phi(\alpha_0 + \alpha_1 X_i)\}]$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

We expect that proportion of fetal malformation can be as low as 1%, and most of fetal weights should be between 0.6 and 1.2, when no EG is administered. These are used for constructing priors for intercept terms α_0 and β_0. We assign normal prior with mean -4.6 (0.01 in logit scale) and relatively large prior variance 10 for α_0. We use a normal prior for β_0 with mean 0.9 (prior guess for average weight) and variance 10. We use noninformative priors, independent mean 0 and variance 10 normal priors, for coefficient parameters α_1, β_1, γ and λ in the three models. As specified in Chapter 3, attenuation parameter for bridge distribution ϕ, positive correlation ρ_1 in Model 2 and correlation coefficient ρ_2 between latent variables Z_{1i} and Z_{2i} have priors $U(0, 1)$, $U(0, 1)$ and $U(-1, 1)$, respectively. Variance σ_r^2 for R_{2i} is assigned $IGa(1, 1)$ distribution, where $IGa(a, b)$ is the inverse gamma density with

shape parameter a and scale parameter b. Variance σ^2 for Y_{2ij} also has $IGa(1,1)$ prior, representing prior belief that weights of fetuses should be less than 2. Priors for parameters are listed below:

Intercept parameters: $\alpha_0 \sim N(-4.6, 10)$; $\beta_0 \sim N(0.9, 10)$;

Coefficient parameters: $\alpha_1 \sim N(0, 10)$; $\beta_1 \sim N(0, 10)$; $\gamma \sim N(0, 10)$; $\lambda \sim N(0, 10)$;

Random effects parameters: $\phi \sim U(0, 1)$; $\rho_1 \sim U(0, 1)$; $\rho_2 \sim U(-1, 1)$;

Variance parameters: $\sigma_r^2 \sim IGa(1, 1)$; $\sigma^2 \sim IGa(1, 1)$.

To check the flexibilities of our informative priors, we conduct a small scale simulation study of prior prediction distributions for fetal malformation probability and fetal weight using Model 3 with dose level 0. The prior predictive distribution is defined as marginal distribution of data, which the integral of the likelihood of the statistical model with respect to the prior distribution:

$$P(\boldsymbol{Y}_{\text{predict}}) = \int L(\boldsymbol{\theta}|Y_{\text{predict}})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

With the specified priors of parameters in Model 3, we generate 10 data sets with 50 clusters and cluster size 1. For each data set, the simulation has the following steps:

1. Draw samples for intercept terms $\alpha_0 \sim N(-4.6, 10)$ and $\beta_0 \sim N(0.9, 10)$;

2. Generate correlated latent variables Z_{1i} and Z_{2i} from bivariate standard normal distribution with correlation ρ_2. Obtain random effects R_{1i} and R_{2i} from generated Z_{1i} and Z_{2i} using the copula model;

3. Sample variances $\sigma_r^2 \sim IGa(1, 1)$ and $\sigma^2 \sim IGa(1, 1)$;

4. Generate error term $\epsilon_{ij} \sim N(0, \sigma^2)$. Obtain one pair of simulated data p_{ij} and Y_{2ij} using (4.1) with dose level 0 and $\gamma = 0$;

5. Repeat step 4 for 50 times.

Figure 4.2 and Figure 4.3 show the box-plots of p_{ij} and Y_{2ij} for the 10 generated data sets. We observe that our priors allow all kinds of predictions for fetal malformation probability and fetal weight with their medians approximately close to our expectation. Hence the informative priors we specified are practical and proper. If we use larger variance priors, we may get unrealistic predictions for Y_{1ij} and Y_{2ij}.

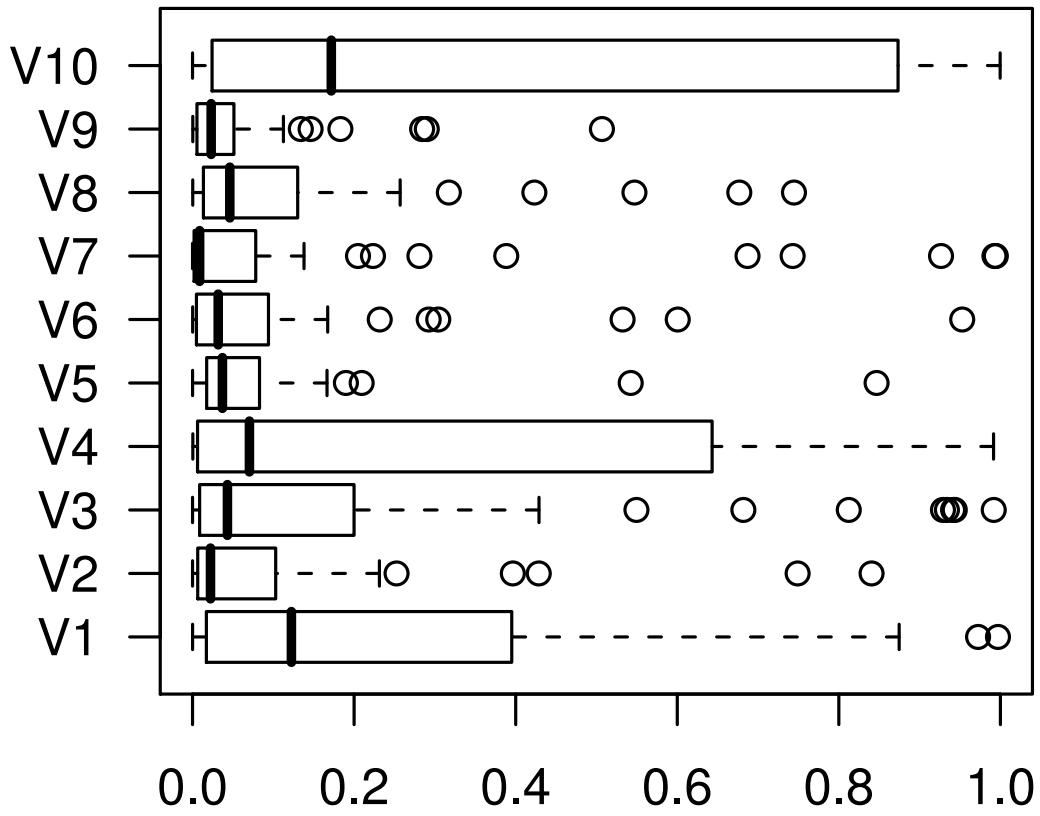

Figure 4.2: Simulated Prior Predictions for Fetal Malformation Probability

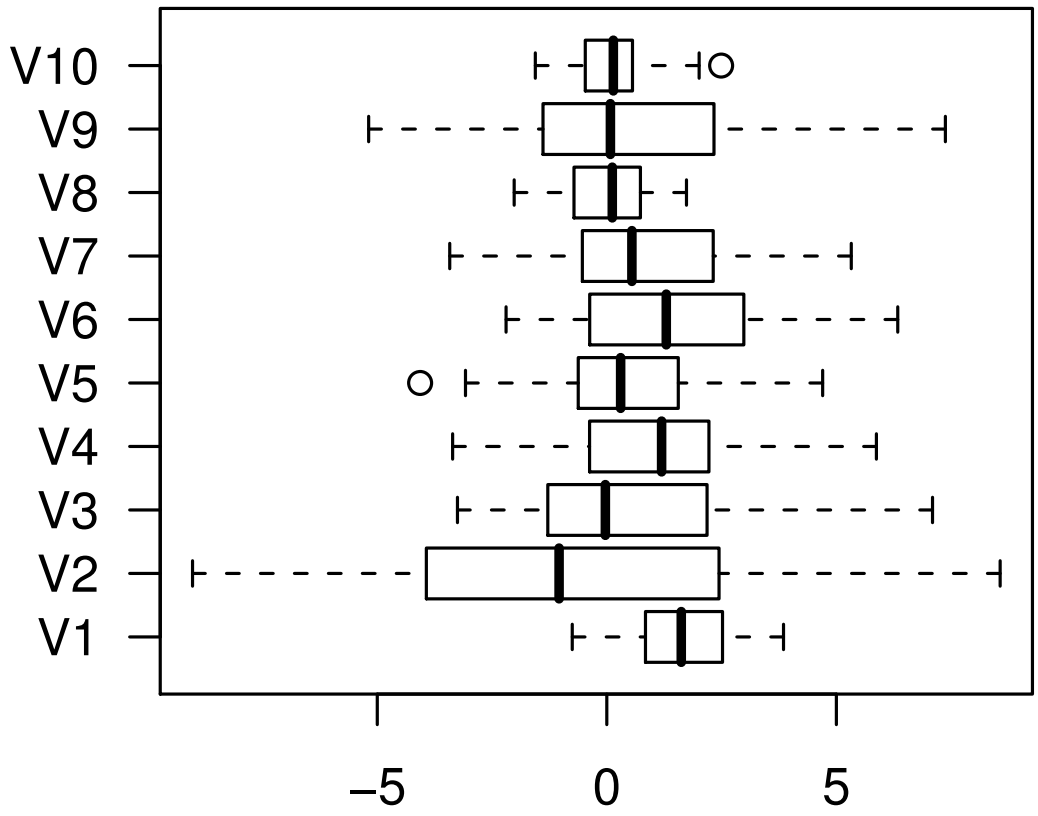

Figure 4.3: Simulated Prior Predictions for Fetal Weight

Table 4.2: Posterior Estimations from Model 1

Parameter	Mean	SD	Median
α_0	-3.947	0.297	-3.932
α_1	1.585	0.153	1.579
β_0	0.960	0.013	0.960
β_1	-0.094	0.008	-0.094
γ	-0.016	0.009	-0.016
λ	-0.074	0.009	-0.074
ϕ	0.787	0.039	0.791
σ	0.088	0.002	0.088

4.3 Posterior Estimates from the Three Models

Samples from posterior distributions of parameters are obtained using MCMC methods implemented via Winbugs [31]. Winbugs is a flexible software for the Bayesian inference of statistical models using Gibbs Sampling method. The Winbugs codes for the parametric Bayesian approach of Model 1, 2, and 3 are given in Appendix B. Sampling methods are used according to the characteristics of the full conditional distributions of parameters which are all continuous. If it is conjugate, standard algorithms could be used for direct sampling; if the conditional distribution is log-concave, derivative-free adaptive rejection sampling method [28] is used; Slice sampling [23] is used for parameters with bounded range and non-log-concave full conditional distribution; for parameters which are not unbounded with non-log-concave conditional distribution, Metropolis-Hastings algorithm [24] [25] is used. In Appendix A, we show that the full conditional distributions of parameters in Model 3 are log-concave or not. If the full conditional distribution is log-concave, then convergence of samples for the corresponding parameter is faster. Slice sampling method is used for ϕ and ρ_2 which have finite supports and non-log-concave conditional distributions. And Metropolis-Hastings algorithm is used for sampling α_0, α_1, R_{1i} $(i = 1, ..., N)$ which have non-log-concave conditional distributions and infinite ranges.

We run three chains each with 10000 Gibbs iterations and randomly generated initial values for the parameters in the three models. Trace plots, autocorrelation plots and Gelman-Rubin diagnostic plots are used to assess convergence of the Gibbs sampler.

Table 4.2, Table 4.3 and Table 4.4 give the posterior estimates of parameters (mean, stan-

Table 4.3: Posterior Estimations from Model 2

Parameter	Mean	SD	Median
α_0	-3.718	0.296	-3.708
α_1	1.396	0.146	1.395
β_0	0.943	0.014	0.943
β_1	-0.079	0.008	-0.079
λ	-0.090	0.007	-0.089
ϕ	0.806	0.040	0.808
ρ_1	0.989	0.010	0.992
σ	0.089	0.002	0.089

Table 4.4: Posterior Estimations from Model 3

Parameter	Mean	SD	Median
α_0	-4.226	0.462	-4.213
α_1	1.720	0.280	1.708
β_0	0.947	0.024	0.950
β_1	-0.087	0.017	-0.087
γ	-0.026	0.009	-0.026
ϕ	0.573	0.056	0.573
ρ_2	-0.813	0.075	-0.828
σ_r	0.172	0.012	0.171
σ	0.087	0.002	0.087

dard deviation, and median) for the two responses from Model 1, 2 and 3 using parametric Bayesian approach. The positive estimates for the dose effect on fetal malformation (α_1) and the negative estimate on fetal weight (β_1) from the three models are consistent with what we observe from Table 4.1: the fetal malformation rate increases and fetal weight decreases as dose level increases.

We observe Model 1 and Model 2 yield similar posterior means (point estimates) for most parameters. Estimate for correlation ρ_1 between Z_{ij} and $Z_{ij'}$ in Model 2 is approximately 1, indicating cluster (litter) has overwhelming bigger effect than subject (fetus) does on associations between the binary and continuous responses within the same subject. Parameter estimates from Model 3 differs from estimations from Model 1 and Model 2, especially for attenuation parameter η (0.57 from Model 3 and 0.79 and 0.81 from Model 1

Table 4.5: CPO Ratios for the Four Bayesian Models

	Model 1/Model 2	Model 3/Model 1	Model 4/Model 3
CPO Ratio	3406.6798	1.2172E+16	52.083

and 2), indicating larger attenuation of conditional parameter estimates of fetal malformation from the correlated random effects model compared to the single random effect models.

Model diagnostics of the three models are implemented based on the posterior predictive distributions of responses. The posterior predictive distribution is an integral of the likelihood with respect to the posterior distribution of parameters defined as:

$$P(\boldsymbol{Y}_{\text{predict}}|\boldsymbol{Y}) = \int L(\boldsymbol{\theta}|Y_{\text{predict}})\pi(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}.$$

For each observation, we calculate the probability mass of fetal malformation using its predictive probability, and the square of estimate error of fetal weight based on its posterior predictive mean. Figure 4.4, Figure 4.5 and Figure 4.6 give the diagnostic plots for Model 1, Model 2 and Model 3 respectively. Predictions of the binary and continuous responses are favorable if the probability mass for the binary response is greater than 0.5 and the square of estimate error for the continuous response is close to 0.

We compare Model 1, 2, and 3 using CPO statistics (3.15) and obtain the CPO ratios for Model 1 to Model 2, Model 2 to Model 3. The CPO ratios for the three models is shown in Table 4.5. From this table, we observe that Model 3 has by far the overwhelming largest CPO value, which implies Model 3 has a better fit of the EG data compared to Model 1 and Model 2.

The estimate of the attenuation parameter ϕ=0.573 from Model 3 confirms that heterogeneity of the clusters causes moderate attenuation of the estimated marginal dose effects on the binary response. The intra-cluster correlation ρ_{Y_1} between the binary responses and ρ_{Y_2} between the continuous responses are estimated by $1-\phi = 1-0.573 = 0.427$ and $\sigma_r^2/(\sigma^2+\sigma_r^2) = 0.796$, respectively. The correlation (ρ_2) between the two cluster-specific random effects R_{1i} and R_{2i}, as well as the within subject dependence coefficient (γ) of Y_{2ij} on Y_{1ij}, both together induces the association between fetal malformation and fetal weight as a whole from the same fetus. The negative values ρ_2=-0.813 and γ=-0.026 suggest negative association between the two responses.

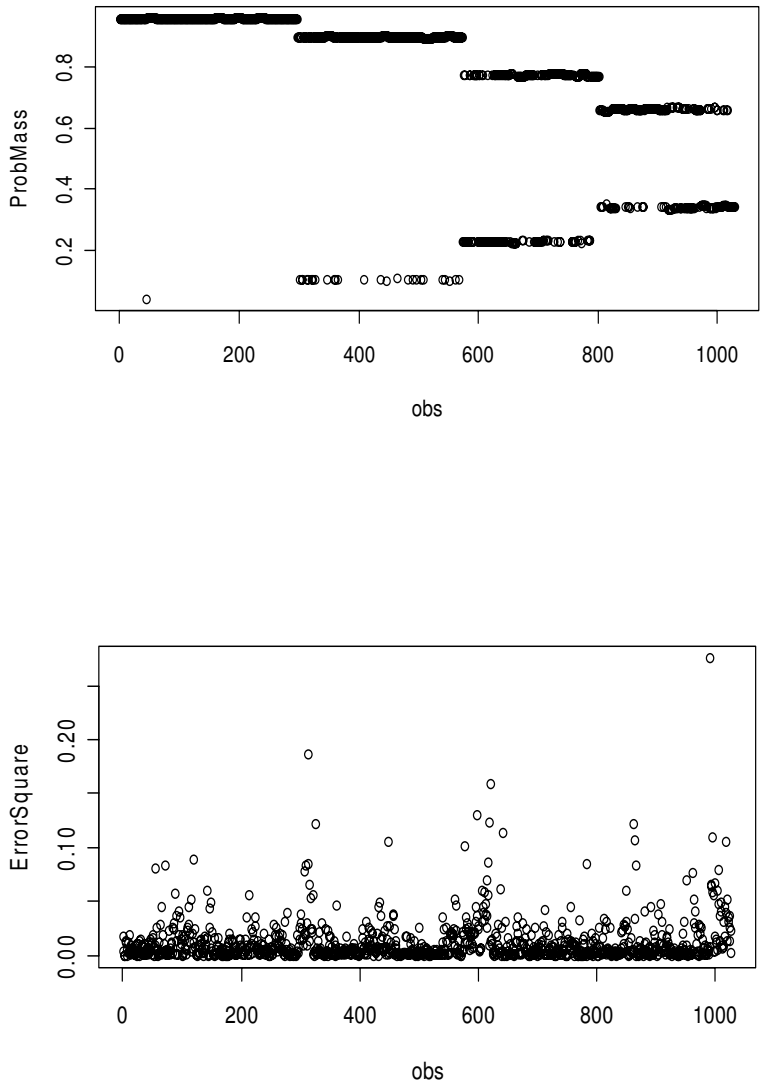

Figure 4.4: Model 1 Diagnostic Plot Based on Posterior Predictive Distributions

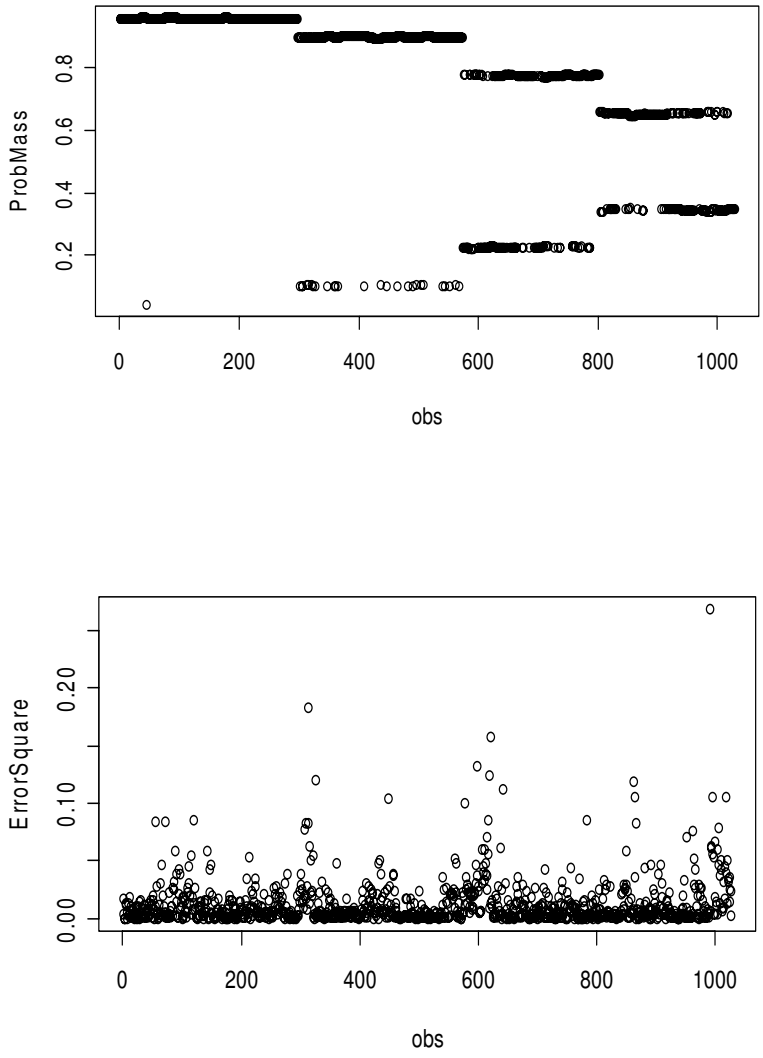

Figure 4.5: Model 2 Diagnostic Plot Based on Posterior Predictive Distributions

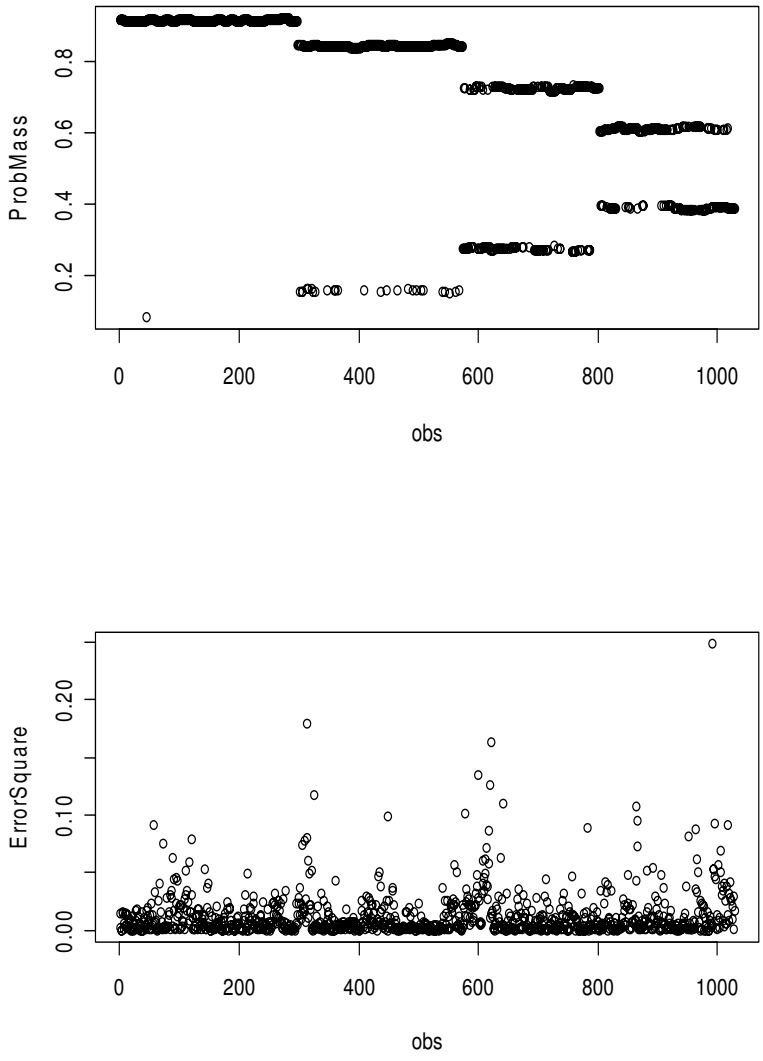

Figure 4.6: Model 3 Diagnostic Plot Based on Posterior Predictive Distributions

CHAPTER 5

SEMIPARAMETRIC BAYESIAN ANALYSIS

In this chapter we present the semiparametric Bayesian method for model analysis using Dirichlet Process prior. The semiparametric Bayesian approach is then applied to the EG data. Comparison of the parametric Bayesian model (Model 3) and semiparametric Bayesian model is based on CPO statistics of the two models.

5.1 Dirichlet Process Prior

Previous proposed models needs the regression error of the continuous response to be normal. The assumption of normality may be an assumption of convenience, but can be restrictive and inappropriate for a study. Now we extend Model 3, which has a better fit than Model 1 and Model 2 for EG data, to a semiparametric model (**Model 4**). This is implemented by the nonparametric mixture modeling via Dirichlet Process (DP) prior. We are concerned about the normality distribution assumption of the error term of the continuous response Y_{2ij}. The Dirichlet Process is used to model the uncertainty about the forms of distribution for the regression error term.

A DP [32], which involves a unknown distribution G, the baseline distribution G_0 as a prior guess of G with $E(G(B)) = G_0(B)$ for any measurable subset B of the support of G_0, and a positive scale precision parameter C, assigns a random probability measures for G on the space of all probability measures. For any measurable partition $(B_1, ..., B_M)$ of the support of G_0, the joint distribution of probabilities $(G(B_1), ..., G(B_M))$ is Dirichlet$(CG_0(B_1), ..., CG_0(B_M))$, where G_0 determines the location of the Dirichlet Process prior, and C characterizes the precision of $G_0(\cdot)$ as the prior guess for $G(\cdot)$.

We keep the regression function for the binary response Y_{1ij} in Model 4 the same as in Model 3. For Y_{2ij} in Model 4, we assume a normal distribution for Y_{2ij} given R_{2i} and

Y_{1ij}, with mean $\mu_{ij} = X_{ij}^T\boldsymbol{\beta} + \gamma(Y_{1ij} - p_{1ij}) + R_{2i}$, and subject-specific variance σ_{ij}^2 (instead of common variance σ^2 as in Model 3), where σ_{ij}^2 for all subjects are assumed to follow independent unknown distribution $G(\cdot)$ with a nonparametric Dirichlet Process prior [13]. In the following we show how this strategy allows us to avoid the normality assumption of the error term from modeling the continuous response Y_{2ij}.

We implement the Dirichlet Process prior using the following hierarchical model

$$Y_{2ij}|R_{2i}, Y_{1ij}; \sigma_{ij}^2 \sim N(\mu_{ij}, \sigma_{ij}^2)$$

with

$$\sigma_{ij}^2|G \sim G(\sigma_{ij}^2)$$

and

$$G(\cdot)|G_0, C \sim DP(G|G_0, C).$$

Then the marginal density of $[Y_{2ij}|R_{2i}, Y_{1ij}; G_0, C]$ has the form

$$E_{DP}\left[\int f_N(Y_{2ij}|\mu_{ij}, \sigma_{ij}^2) dG(\sigma_{ij}^2)|G_0, C\right],$$

where $f_N(Y_{2ij}|\mu_{ij}, \sigma_{ij}^2)$ is the $N(\mu_{ij}, \sigma_{ij}^2)$ density and E_{DP} is the expectation taken with respect to the DP prior of $G(\sigma_{ij}^2)$ given (G_0, C). Thus marginal distribution of $[Y_{2ij}|R_{2i}, Y_{1ij}; G_0, C]$ has a density of Dirichlet mixture of normal. In this way, we relax the normality assumption of regression error of Y_{2ij} with a large semiparametric class of symmetric and unimodal densities.

5.2 Prior and Posterior

Hyperparameter C of the unknown distribution G plays an important role in the semiparametric Bayesian approach for our models. When C is large, distribution G has many support points. And G becomes the parametric distribution G_0 if C approaches infinity. If C is small, then G would concentrate only on a few points. There exist a neat data augmentation algorithm to generate samples of C if C is assigned a prior of gamma distribution [33]. Assume $C \sim Ga(a, b)$ with $a > 0$ as the shape parameter and $b > 0$ as the scale parameter for the gamma distribution, then this prior gives mass to both large and small values for C if a is small and b is large, e.g. $C \sim Ga(0.01, 100)$ with mean 1 and variance 100 [13], which

implies that there is reasonable possibility that the unknown distribution G is either close to G_0 or it is only concentrated on a few atoms.

The prior guess G_0 of G is assumed to be inverse gamma. We keep the priors $\pi(\boldsymbol{\theta})$ for other parameters the same as they are in the parametric method of Model 3. The joint posterior distribution of Model 4 is then proportional to

$$\left[\prod_{i=1}^{N}\left\{\prod_{j=1}^{n_i} L(\boldsymbol{\theta_1}, \sigma_{ij}^2, R_{1i}, R_{2i}|Y_{1ij}, Y_{2ij})dG(\sigma_{ij}^2)\right\} f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2})\right] \pi(\boldsymbol{\theta})DP(G|G_0, C)\pi(C),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ with $\boldsymbol{\theta_1} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ and $\boldsymbol{\theta_2} = (\phi, \rho_2, \sigma_r^2)$, and $\pi(C)$ is the gamma density. The posterior distribution is again difficult for us to evaluate analytically, however, it is possible to sample from it using MCMC tools.

5.3 Application to EG Data

To apply the semiparametric model (Model 4) to the EG data, we assign the precision parameter C a $Ga(1, 10)$ prior, which has mean 10 and variance 100. Hence our prior put reasonable mass on both high and low values for C. And the baseline distribution G_0 has $IGa(1, 1)$ prior distribution. We keep priors of other parameters the same as priors in the parametric approach for Model 3.

The Dirichlet Process prior is implemented in Winbugs [31] using stick-breaking construction proposed by Sethuraman in 1994 [12] [19]. For each iteration in MCMC and a fixed positive integer M, we first draw M samples $(\nu_1, ..., \nu_M)$ from the baseline distribution $IGa(1, 1)$. Let random vector $(\delta_1, ..., \delta_{M-1})$ be independently and identically distributed with beta distribution $Be(1, C)$ determined by precision parameter C, and $\delta_M = 1$. Then mixture weight p_m $(m = 1, ..., M)$ is constructed by stick-breaking procedure:

$$p_1 = \delta_1,$$

$$p_2 = \delta_2(1 - \delta_1),$$

$$p_3 = \delta_3(1 - \delta_2)(1 - \delta_1),$$

$$\cdots\cdots$$

$$p_M = \delta_M(1 - \delta_{M-1}) \cdots (1 - \delta_1).$$

We have

$$\sum_{1\leq m\leq M} p_m = 1 - \prod_{1\leq m\leq M} (1-\delta_m) \to 1$$

as $M \to \infty$. We select the most appropriate ν_m as the value for the subject-specific variance σ^2_{ij} where the index m is generated from categorical distribution with probabilities $(p_1, ..., p_M)$ for integers $(1, ..., M)$. The corresponding Winbugs code for the semiparametric Bayesian approach is given in Appendix B. We again run three chains each with 10000 Gibbs iterations and randomly generated initial values for the parameters in Model 4. Convergence of the Gibbs sampler is checked by trace plots, autocorrelation plots and Gelman-Rubin diagnostic plots.

Posterior estimates of parameters from Model 4 is shown in Table 5.1. Both parametric method for Model 3 and semiparametric method for Model 4 yield similar posterior means for regression parameters α_0, α_1, β_0, β_1, and γ. However, we obtain larger estimate for attenuation parameter ϕ, smaller estimate for correlation parameter ρ_2 and smaller estimate for between-cluster variance σ^2_r from Model 4 compared to Model 3 (ϕ : 0.573 from Model 3 and 0.693 from Model 4; ρ_2 : -0.813 from Model 3 and -0.573 from Model 4; σ^2_r : 0.172 from Model 3 and 0.086 from Model 4). This may be due to the relaxation of normality assumption for the continuous response in semiparametric model, which would allow heavier tail for the regression error of Y_{2ij}. Hence part of the intra-cluster variability in Model 3 are considered as the error of modeling Y_{2ij} in Model 4. Thus less cluster heterogeneity, less within-cluster association and less between-cluster variance are obtained from the modeling results of Model 4 compared to Model 3.

For almost all the parameters, posterior standard deviations from parametric method of Model 3 are larger than posterior standard deviations from semiparametric method of Model 4. These are contrary to our expectations that posterior standard deviations from semiparametric model should be larger than the corresponding posterior standard deviation from parametric model when the parametric assumption is correct. This indicates that the semiparametric model of Model 4 fits the EG data better compared to the parametric model of Model 3.

We also compare Model 3 and Model 4 using CPO statistics. From Table 4.5, we obtain that $CPO_{M_3}/CPO_{M_4} = 0.0192 < 1$, which again implies the EG data is better fitted by the semiparametric model 4 than the parametric model 3.

Table 5.1: Posterior Statistics from Model 4

Parameter	Mean	SD	Median
α_0	-4.257	0.347	-4.246
α_1	1.721	0.196	1.710
β_0	0.950	0.014	0.950
β_1	-0.086	0.008	-0.086
γ	-0.032	0.009	-0.032
ϕ	0.693	0.048	0.695
ρ_2	-0.573	0.109	-0.583
σ_r	0.086	0.007	0.086
C	21.220	12.820	18.690

CHAPTER 6

MAXIMUM LIKELIHOOD METHOD

Model 3 is applied to the EG data using likelihood method in this chapter. We also use fractional polynomial regression to improve modeling results. Simulation studies are conducted using ML method to investigate the robustness of Model 3.

6.1 Data Application Using MLE

The full likelihood of Model 3 is given in (3.12). As discussed earlier, unlike CR and FL model, our likelihood is tractable and maximization is straightforward. In order to use the ML method for parameter estimation, the likelihood is obtained via taking the expectation with respect to the joint density of (R_{1i}, R_{2i}). These integral computations can be implemented using nonadaptive Gaussian quadrature techniques in PROC NLMIXED in SAS (v9.1) software along with optimization techniques such as Newton-Raphson with ridging. (SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.) The SAS code of Model 3 is given in Appendix B.

Table 6.1 shows the ML parameter estimations from Model 3. P-values for parameters are all less than 0.001, indicating all of them are significant. Marginal regression parameter estimates for the binary response Y_{1ij} and continuous response Y_{2ij} have slightly different point estimates compared to the FL model (Table 2 in Fitzmaurice and Laird, 1995 [6]). But estimated standard errors of the parameter estimates are smaller than those of the FL model, especially for the binary response (approximately 20%), leading to tighter 95% confidence intervals. The intra-cluster correlation $\rho_{Y_1} = 0.338$ between the binary responses and $\rho_{Y_2} = 0.626$ between the continuous responses are somewhat greater than the estimates of the intracluster correlations obtained from the FL model.

Table 6.1: ML Estimates of Model 3

Parameter	Estimate	SE	p-value
	Fetal Malformation		
α_0	-4.493	0.439	<0.001
α_1	1.758	0.204	<0.001
	Fetal Weight		
β_0	0.953	0.012	<0.001
β_1	-0.084	0.006	<0.001
γ	-0.034	0.008	<0.001
	Other Parameters		
ϕ	0.662	0.041	<0.001
σ_r	0.095	0.005	<0.001
σ	0.074	0.002	<0.001
ρ_{Y_1}	0.338	0.041	<0.001
ρ_{Y_2}	0.626	0.029	<0.001
ρ_2	-0.643	0.064	<0.001

Using ML estimates from Model 3 and expressions (2.12)-(2.15), we obtain from a Monte Carlo simulation that, for dose level 0, the estimated correlation between the binary and continuous responses from the same fetus is $\text{Corr}[Y_{1ij}, Y_{2ij}] = -0.1481$, and the estimated correlation between the two responses from different fetuses within the same cluster is $\text{Corr}[Y_{1ij}, Y_{2ij'}] = -0.1245$. This shows that the former correlation is approximately 19% higher than latter correlation as expected.

In the smoothed residual plots of fitting results from Model 3 using ML method, obtained via the LOESS procedure in SAS [34], we observe some quadratic trends of covariate dose for both the fetal malformation and fetal weight responses, as shown in Figure 6.1 and Figure 6.2. Thus we re-analyze the EG data by adding quadratic terms of dose effect for both responses in **Model 5**:

$$\begin{aligned} &\text{logit}(E[Y_{1ij}|R_{1i}, X_i]) = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + R_{1i} \\ &Y_{2ij} = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \gamma(Y_{1ij} - p_{1i}) + \lambda R_{2i} + \epsilon_{ij}. \end{aligned} \tag{6.1}$$

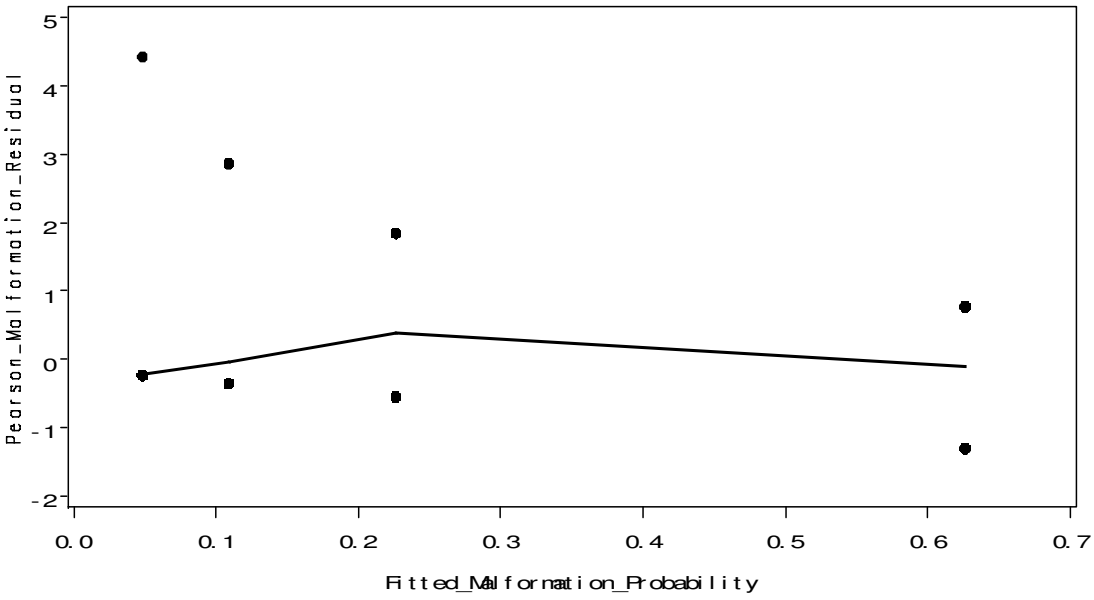

Figure 6.1: Residual Plot for Fetal Malformation of Model 3

No further trends are detected from corresponding residual plots for the two responses, as shown in Figure 6.3 and Figure 6.4. Estimates of the parameters in Model 5 are given in Table 6.2, where α_2 and β_2 are the parameters of squared dose effect for the two responses. P-values for α_2 and β_2 are <0.001 and 0.001, respectively. P-values for all the other parameters are less than 0.001. This suggests that the quadratic dose effect terms are significant and necessary.

6.2 Alternative Models

For comparison with results from Model 3, we consider a more complex correlation structure. We change the parameter γ of Model 3 to $(\gamma_1 + \gamma_2 X_i)$ to verify whether the dose effect on

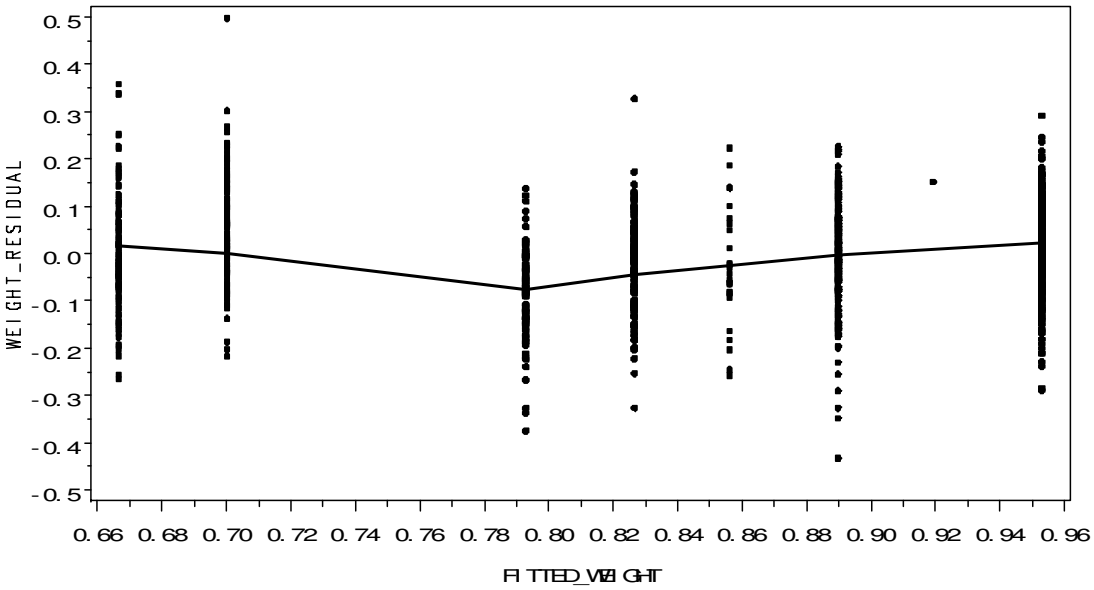

Figure 6.2: Residual Plot for Fetal Weight of Model 3

fetal weight is the same for both the normal and malformed fetuses. However, we obtain the P-value for γ_2 is 0.7154, which indicates that the dose effect on fetal weight does not depend malformation.

Our method allows use of fractional polynomial regression equations [9] for $E(Y_{1ij}|X_i)$ and $E(Y_{2ij}|X_i)$. To model the relationship of dose with both responses, we have fitted several fractional polynomial models including models using $(X_i, \sqrt{X_i})$ and $(1/X_i, 1/\sqrt{X_i})$ to the development toxicity data:

$$\text{logit}(E[Y_{1ij}|R_{1i}, X_i]) = \alpha_0 + \alpha_1 X_i + \alpha_2 \sqrt{X_i} + R_{1i}$$
$$Y_{2ij} = \beta_0 + \beta_1 X_i + \beta_2 \sqrt{X_i} + \gamma(Y_{1ij} - p_{1i}) + \lambda R_{2i} + \epsilon_{ij},$$

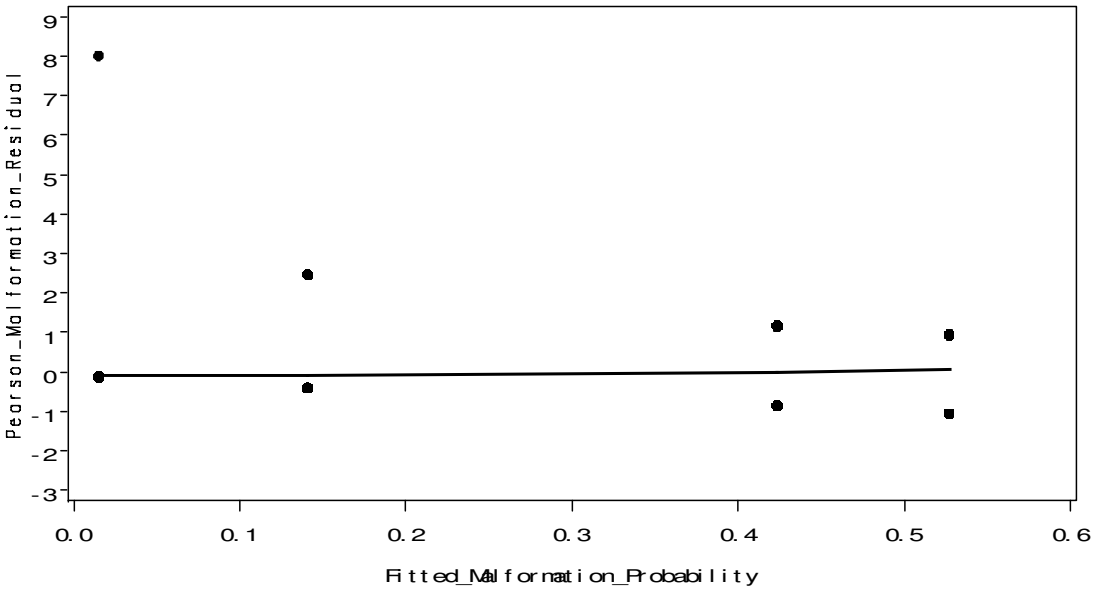

Figure 6.3: Residual Plot for Fetal Malformation of Model 5

and

$$\text{logit}(E[Y_{1ij}|R_{1i}, X_i]) = \alpha_0 + \alpha_1 \frac{1}{X_i} + \alpha_2 \frac{1}{\sqrt{X_i}} + R_{1i}$$

$$Y_{2ij} = \beta_0 + \beta_1 \frac{1}{X_i} + \beta_2 \frac{1}{\sqrt{X_i}} + \gamma(Y_{1ij} - p_{1i}) + \lambda R_{2i} + \epsilon_{ij},$$

The results from these two models are shown in Table 6.3 and Table 6.4. We compare these two models with Model 3 and Model 5 using Akaike's Information Criterion (AIC) [35] and Bayesian Information Criterion (BIC) [36] values. AIC is a measure of the goodness of fit of a statistical model, developed by Akaike. It is defined as

$$AIC = 2k - 2\ln(L),$$

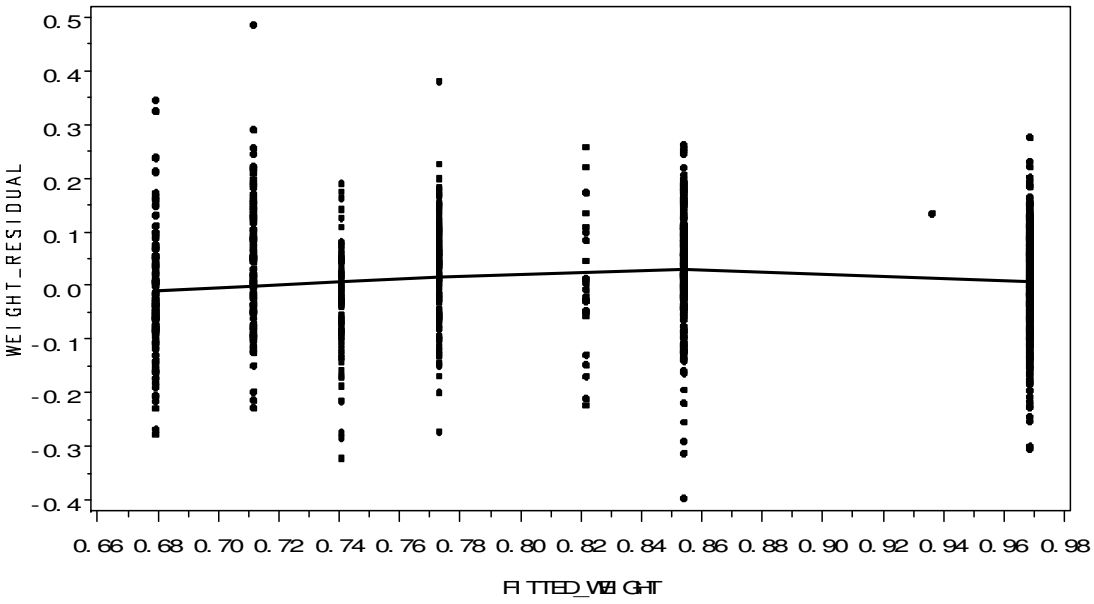

Figure 6.4: Residual Plot for Fetal Weight of Model 5

where k is the number of free parameters to be estimated and L is the maximized value of the likelihood for the estimated model. Schwarz proposed BIC with the fomula

$$BIC = k\ln(n) - 2\ln(L),$$

where n is the number of observations. BIC penalizes free parameters more strongly than does AIC. Both AIC and BIC are tools for model selection. The model with the lowest AIC (or BIC) value is ranked as the best model among several competing models for a given data set.

The corresponding AIC and BIC values of model with covariates $(X_i, \sqrt{X_i})$, model with covariates $(1/X_i, 1/\sqrt{X_i})$ and Model 3, are bigger than the AIC and BIC of Model 5, as given in Table 6.5, suggesting Model 5 has by far the best fit of EG data. For experiments

Table 6.2: ML Estimates of Model 5 with Quadratic Dose Effects

Parameter	Estimate	SE	p-value
	Fetal Malformation		
α_0	-5.735	0.717	<0.001
α_1	5.120	1.003	<0.001
α_2	-1.053	0.268	<0.001
	Fetal Weight		
β_0	0.969	0.012	<0.001
β_1	-0.175	0.027	<0.001
β_2	0.030	0.009	0.001
γ	-0.032	0.008	<0.001
	Association Parameters		
ϕ	0.726	0.052	<0.001
σ_r	0.086	0.007	<0.001
σ	0.074	0.002	<0.001
ρ_{Y_1}	0.274	0.052	<0.001
ρ_{Y_2}	0.577	0.039	<0.001
ρ_2	-0.611	0.092	<0.001

with more dose levels than we have, such a fractional polynomial method may become more useful to model the dose-response effects.

To see the impact of associations between fetal malformation Y_{1ij} and fetal weight Y_{2ij} on the estimates of the regression parameters, we fit regression models assuming of no correlation between R_{1i} and R_{2i} as well as no dependence of Y_{2ij} on Y_{1ij}, i.e. $\rho_2 = 0$ and $\gamma = 0$ (**Model 3a**). We consider **Model 3b** with $\gamma = 0$ to accommodate the special case no direct within-subject association between Y_{1ij} and Y_{2ij}. We also consider separate and independent regression model ($\rho_2 = 0$ and $\gamma = 0$) for Y_{1ij} and Y_{2ij} with quadratic covariate ${X_i}^2$ added (**Model 5a**).

The regression results of these three models are given in Table 6.6. No noticeable changes in point estimates are observed comparing fitting results of Model 3a and Model 3b with Model 3, and Model 5a with Model 5. However, we observe the estimated standard errors for parameters from Model 3a and Model 3b are larger compared to corresponding estimates

Table 6.3: ML Estimates of Model with Covariates $(X_i, \sqrt{X_i})$

Parameter	Estimate	SE	p-value
	Fetal Malformation		
α_0	-6.552	1.122	<0.001
α_1	-0.727	0.778	0.353
α_2	-5.221	1.877	0.007
	Fetal Weight		
β_0	0.965	0.012	<0.001
β_1	-0.061	0.010	<0.001
β_2	-0.034	0.008	<0.001
γ	-0.087	<0.001	<0.001
	Association Parameters		
ϕ	0.751	0.043	<0.001
σ_r	0.089	0.007	<0.001
σ	0.074	0.002	<0.001
ρ_{Y_1}	0.249	0.043	<0.001
ρ_{Y_2}	0.590	0.042	<0.001
ρ_2	-0.629	0.081	<0.001

from Model 3. This indicates that substantially larger sample sizes are required while using Model 3a and Model 3b in order to obtain similar standard errors as Model 3. Similarly, larger standard errors are observed when comparing results for the regression parameters of the two responses from Model 5a to Model 5. Particularly for this data set, the p-values for the association parameter (ρ_2) between the two random effects and the direct dependence (γ) of Y_{2ij} on Y_{1ij} are both less than 0.001 in Model 3 and Model 5, implying significant association between the two responses. AIC and BIC of the five fitted models are shown in Table 6.7. Model 5 has by far the smallest AIC and BIC values (with differences larger than 10 compared to the AIC and BIC obtained from Model 3), which indicates it is the best fitted model for our data set.

Table 6.4: ML Estimates of Model with Covariates $(1/X_i, 1/\sqrt{X_i})$

Parameter	Estimate	SE	p-value
	Fetal Malformation		
α_0	1.423	0.669	0.036
α_1	0.225	0.081	0.007
α_2	-3.086	0.856	0.001
	Fetal Weight		
β_0	0.834	0.012	<0.001
β_1	0.004	0.001	<0.001
β_2	-0.029	0.008	<0.001
γ	-0.082	<0.001	<0.001
	Association Parameters		
ϕ	0.719	0.045	<0.001
σ_r	0.109	0.007	<0.001
σ	0.074	0.002	<0.001
ρ_{Y_1}	0.281	0.045	<0.001
ρ_{Y_2}	0.685	0.030	<0.001
ρ_2	-0.551	0.062	<0.001

6.3 Simulation Studies of Robustness

In theory, the only advantage GEE has over our full-likelihood approach is when the marginal regression model is correctly specified, but the full likelihood is mis-specified; in this case, the GEE estimate of the marginal regression parameters will be asymptotically unbiased, but those from our full likelihood could be biased.

To investigate the robustness of ML approach of Model 3, we performed two simulation studies under misspecification of regression function of the continuous response as well as misspecification of the random effect distribution for the binary response. Our current model (Model 3) assumes that the parameter (γ) measuring the direct association between the two responses within the same subject does not depend on dose (or other covariates). Also Model 3 assumes the random effect for the binary response has a bridge distribution for the logit link function.

Table 6.5: Fit Statistics for Fractional Polynomial Models

Fit Statistics	Model 3	Model 5	Model with $(X_i, \sqrt{X_i})$	Model $(1/X_i, 1/\sqrt{X_i})$
-2 Log Likelihood	-1311	-1331	-1313	-1284
AIC	-1293	-1309	-1291	-1262
BIC	-1271	-1281	-1263	-1234

Table 6.6: ML Estimates of Parameters from Model 3a, Model 3b and Model 5a

	Model 3a	Model 3b	Model 5a
Parameter	Estimate (SE)	Estimate (SE)	Estimate (SE)
		Fetal Malformation	
α_0	-4.183 (0.453)	-4.489 (0.410)	-5.971 (0.800)
α_1	1.834 (0.250)	1.887 (0.215)	5.321 (1.126)
α_2			-1.072 (0.300)
		Fetal Weight	
β_0	0.957 (0.017)	0.954 (0.013)	0.986 (0.013)
β_1	-0.095 (0.007)	-0.095 (0.006)	-0.166 (0.028)
β_2			0.024 (0.009)
		Other Parameters	
ϕ	0.725 (0.044)	0.732 (0.042)	0.749 (0.052)
σ_r	0.091 (0.008)	0.088 (0.005)	0.085 (0.010)
σ	0.075 (0.002)	0.074 (0.002)	0.075 (0.002)
ρ_{Y_1}	0.276 (0.044)	0.268 (0.042)	0.251 (0.052)
ρ_{Y_2}	0.597 (0.046)	0.581 (0.032)	0.562 (0.059)
ρ_2		-0.700 (0.057)	

Table 6.7: Fit Statistics for Five Models

Fit Statistics	Model 3a	Model 3b	Model 5a	Model 3	Model 5
-2 Log Likelihood	-1254	-1278	-1292	-1311	-1331
AIC	-1240	-1262	-1274	-1293	-1309
BIC	-1223	-1242	-1251	-1271	-1281

Table 6.8: Parameter Values for Simulation Studies of Robustness

α_0	α_1	β_0	β_1	γ ($\gamma_1 = \gamma_2$)	ϕ	σ_r	σ	ρ_2
-4.493	1.758	0.953	- 0.084	0.034	0.662	0.095	0.074	-0.643

Table 6.9: Simulation Results of Regression Function Misspecification of Model 3

	Dose Effect (Malformation)		Dose Effect (Weight)	
	Mean S.E.	Rel. Bias	Mean S.E.	Rel. Bias
FL Model	0.157	0.102	0.008	0.092
Model 3	0.136	0.106	0.009	0.102

In the first simulation study, we generate 50 data sets from a modification of Model 3 with γ replaced by $\gamma_1 + \gamma_2 X_i$, while assuming $\gamma_1 = \gamma_2$ for simplicity. Each data set contains 100 clusters with fixed cluster size 10. And values used for parameters to generate such data set is given in Table 6.8.

Both our likelihood approach of Model 3 and FL model using a GEE method (one-step) are applied to the simulated 50 data sets. The mean standard errors and relative biases for estimates of the marginal dose effect $\phi\alpha_1$ and β_1 are shown in Table 6.9, as we are concerned about the bias of regression estimates. From Table 6.9 we observe that our model is robust to regression model misspecifications compared to FL approach.

We also generate data from a modification of Model 3, where distribution of the random effect R_{1i} of the binary response is assumed to be standard normal instead of bridge. We use the value shown in Table 6.8 for each parameter. Again 50 data sets are simulated and each set contains 100 clusters with fixed cluster size 10. The average standard errors and relative biases of the estimated dose effects on fetal malformation (conditional) α_1 and fetal weight (marginal) β_1 from fitting results of Model 3 are given in Table 6.10. Marginal estimates is not easy to obtain for the binary response in this case, since after integrated out R_{1i} which has standard normal distribution, the marginal function form of Y_{1ij} usually dose not preserve logistic.

In this small scale simulation, we observe that our full likelihood approach is also robust to random effects distribution misspecification as far as the bias of regression estimate is

Table 6.10: Simulation Results of Random Effect Distribution Misspecification of Model 3

	Dose Effect (Malformation)		Dose Effect (Weight)	
	Mean S.E.	Rel. Bias	Mean S.E.	Rel. Bias
Model 3	0.136	0.123	0.009	0.009

concerned. This makes our approach an extremely attractive alternative to the GEE based FL approach.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this dissertation, we have developed two single random effect models as well as a correlated bivariate random effects model for bivariate discrete and continuous outcomes motivated by an application in developmental toxicology study, where four different dose levels of EG are administered to 94 pregnant mice and fetal malformation and weight are measured for each fetus from the 94 clusters. Our models are attractive in that it provides both conditional and marginal interpretation of the dose-response modeling which happens to be the cornerstone of quantitative risk assessment. Similar to a generalized linear mixed model (GLMM) setup, the binary responses (fetal malformation) are related to the covariates through a logit link whereas the continuous responses (fetal weight) are related to the covariates by an identity link. Our models can efficiently handle the effects of clustering (litter effect) through introduction of cluster-specific random effects in a regression setup, as well as the association between outcomes (fetal weight and malformations) on the same (different) fetus.

In the first random effect model, a single cluster-specific random effect is used for both binary and continuous responses to characterize the association between the two outcomes due to cluster effect. Model for the continuous response is conditioned on the binary response so that the association between the mixed responses within the same subject induced by subject effect is modeled by the direct dependence of the continuous response on the binary response. Similarly, direct effect of the binary response on the continuous response is used to characterize subject effect in Model 3; Meanwhile, two separate but correlated cluster random effects are used for cluster effect in this model. On the contrary, a single subject-specific random effect for both responses is used in the Model 2 for the association between the two responses induced by both cluster and subject effects.

As Bayesian method has several advantages over the frequentist method, both parametric

and semiparametric Bayesian approaches are applied to the three random effects model for statistical inference. We incorporate extra information in prior elicitation for model efficiency. In parametric Bayesian analysis of the three models, we assume a normal distribution with constant variance (with inverse gamma prior) for error terms in the regression model of the continuous response. Model selection among the single random effect models and the correlated bivariate random effects model is conducted based on CPO statistics of these models.

In the semiparametric Bayesian analysis, we extend the correlated bivariate random effects model, which fits the motivation data better than the other two models, to a semiparametric model via assuming the regression error of the continuous response to have subject-specific variance with a unknown distribution. And this unknown distribution has a nonparametric Dirichlet Process prior assigned. Hence, the continuous response given the binary response and the cluster-specific random effect has a Dirichlet mixture of normal distribution. We select the semiparametric model for its higher CPO value and smaller posterior standard deviations of parameters compared to parametric correlated bivariate random effects model. This suggests that when appropriate, the semiparametric Bayesian analysis produces a more precise estimates of parameters than those from parametric analysis in this kind of study.

We then propose Maximum Likelihood (ML) method for model analysis. With a correctly specified model, provided estimation techniques are computationally feasible, the ML method reigns supreme over any other estimation methods for its fidelity to the likelihood principle as well as for its asymptotic properties. There are clearly several advantages of our ML method over the GEE method used in the FL approach. Firstly, although ML estimation appears to be complicated in a clustered mixed responses setting, our model is easily amenable to ML estimation through routine optimization techniques readily available in standard software like SAS (v9.1). Secondly, ML estimates are asymptotically efficient [37]. Thus, using our method we achieve smaller standard errors of the parameter estimates and consequently tighter confidence intervals over the estimates obtained by the GEE method. Thirdly, GEE methods provide only marginal models, which is unable to measure the attenuation of the dose effect to the population due to the heterogeneity induced by clustering (litter effect). Our model has both conditional and marginal interpretation in this context. Robustness of Model 3 to regression function misspecifications as well as random effect distribution

misspecifications using ML method is investigated by two simulation studies.

de Leon and Carriere [4] developed models for the joint distribution of nominal, ordinal and continuous variables with parameter estimates based on the maximization of the full likelihood. One of our interesting future work would be to extend our existing models to a trivariate model to joint model clustered data with binary, continuous and count responses. Bridge distribution would again be used for the random effect of the binary response so that both marginal and conditional models could be physically interpreted for the binary response.

APPENDIX A

PROOF OF LOG-CONCAVITY OF FULL CONDITIONAL DISTRIBUTIONS

In the following, we investigate the log-concavity properties of the full conditional distributions of regression parameters as well as random effects R_{1i} and R_{2i} $(i = 1, ..., N)$ in Model 3.

We have $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \sigma^2)$ with $\boldsymbol{\theta_1} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ and $\boldsymbol{\theta_2} = (\phi, \rho_2, \sigma_r^2)$ in Model 3, where $\boldsymbol{\alpha} = (\alpha_0, ..., \alpha_{K-1})$ with normal prior $\alpha_k \sim N(\mu_{\alpha_k}, \sigma_{\alpha_k}^2)$ $(k = 0, ..., K-1)$, $\boldsymbol{\beta} = (\beta_0, ..., \beta_{K-1})$ with normal prior $\beta_k \sim N(\mu_{\beta_k}, \sigma_{\beta_k}^2)$ $(k = 0, ..., K-1)$, $\gamma \sim N(\mu_\gamma, \sigma_\gamma^2)$, bridge distribution parameter $\phi \sim U(0, 1)$, correlation parameter $\rho_2 \sim U(-1, 1)$, precision parameters $1/\sigma_r^2 \sim Ga(a, b)$, and $1/\sigma^2 \sim Ga(c, d)$. Here μ_{α_k}, μ_{β_k}, μ_γ, σ_{α_k}, σ_{β_k}, σ_γ, a, b, c and d are known values.

Let $\boldsymbol{\theta}_{-\omega}$ denote the parameter vector $\boldsymbol{\theta}$ with ω excluded, where ω is the random variable and $\boldsymbol{\theta}_{-\omega}$ is held fixed, and $X_{ij} = (x_{ij0}, x_{ij1}, ..., x_{ij(K-1)}) = (1, x_{ij1}, ..., x_{ij(K-1)})$.

α_k $(k = 0, ..., K-1)$:

The full conditional distribution of α_k is:

$$
\begin{aligned}
&\pi(\alpha_k|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{\theta}_{-\alpha_k}) \\
\propto &L(\alpha_k, \boldsymbol{\theta}_{-\alpha_k}, \boldsymbol{R_1}, \boldsymbol{R_2}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2}) \pi(\alpha_k, \boldsymbol{\theta}_{-\alpha_k}) \\
\propto &\left[\prod_{i=1}^{N} \prod_{j=1}^{n_i} \frac{\exp\{(x_{ij}^T \boldsymbol{\alpha} + r_{1i}) y_{1ij}\}}{1 + \exp(x_{ij}^T \boldsymbol{\alpha} + r_{1i})} \exp\left[\frac{-\{y_{2ij} - x_{ij}^T \boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}^2}{2\sigma^2} \right] \right] \\
&\times \exp\left\{ \frac{-(\alpha_k - \mu_{\alpha_k})^2}{2\sigma_{\alpha_k}^2} \right\},
\end{aligned}
$$

where $p_{1ij} = \exp(\phi x_{ij}^T \boldsymbol{\alpha}) / \{1 + \exp(\phi x_{ij}^T \boldsymbol{\alpha})\}$. Hence kernel of the logarithm of the conditional

distribution is

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[(x_{ij}^T\boldsymbol{\alpha}+r_{1i})y_{1ij}-\log\{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}\right]$$
$$-\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{\{y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma(y_{1ij}-p_{1ij})-r_{2i}\}^2}{2\sigma^2}\right]-\frac{(\alpha_k-\mu_{\alpha_k})^2}{2\sigma_{\alpha_k}^2}.$$

Then

$$\begin{aligned}
&\frac{d\log\pi(\alpha_k|\boldsymbol{Y_1},\boldsymbol{Y_2},\boldsymbol{\theta}_{-\alpha_k})}{d\alpha_k}\\
=&\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[x_{ijk}y_{1ij}-\frac{\{\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}x_{ijk}}{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})}\right]-\frac{\alpha_k-\mu_{\alpha_k}}{\sigma_{\alpha_k}^2}\\
&-\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{\gamma}{\sigma^2}\left[y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma\left\{y_{1ij}-\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})}{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})}\right\}-r_{2i}\right]\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})\phi x_{ijk}}{\{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})\}^2}\right],
\end{aligned}$$

and

$$\begin{aligned}
&\frac{d^2\log\pi(\alpha_k|\boldsymbol{Y_1},\boldsymbol{Y_2},\boldsymbol{\theta}_{-\alpha_k})}{d\alpha_k^2}\\
=&-\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{\{\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}x_{ijk}^2}{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})}-\frac{\{\exp(2x_{ij}^T\boldsymbol{\alpha}+2r_{1i})\}x_{ijk}^2}{\{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}^2}\right]-\frac{1}{\sigma_{\alpha_k}^2}\\
&-\frac{\gamma\phi x_{ijk}}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\exp(2\phi x_{ij}^T\boldsymbol{\alpha})}{\{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})\}^4}\gamma\phi x_{ijk}\\
&-\frac{\gamma\phi x_{ijk}}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma\left\{y_{1ij}-\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})}{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})}\right\}-r_{2i}\right]\\
&\times\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})\phi x_{ijk}-\exp(2\phi x_{ij}^T\boldsymbol{\alpha})\phi x_{ijk}}{\{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})\}^3}\\
=&-\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})}{\{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}^2}x_{ijk}^2\right]-\frac{1}{\sigma_{\alpha_k}^2}-\frac{\gamma^2\phi^2x_{ijk}^2}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\exp(2\phi x_{ij}^T\boldsymbol{\alpha})}{\{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})\}^4}\\
&-\frac{\gamma\phi^2x_{ijk}^2}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma\left\{y_{1ij}-\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})}{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})}\right\}-r_{2i}\right]\\
&\times\frac{\exp(\phi x_{ij}^T\boldsymbol{\alpha})-\exp(2\phi x_{ij}^T\boldsymbol{\alpha})}{\{1+\exp(\phi x_{ij}^T\boldsymbol{\alpha})\}^3}.
\end{aligned}\tag{A.1}$$

The first three terms in (A.1) are negative, however, the last term is not always negative, hence the full conditional distribution of α_k $(k=0,...,K-1)$ is not log-concave and

Metropolis-Hastings algorithm is used for sampling α_k which has infinite range.

β_k $(k = 0, ..., K-1)$:

The full conditional distribution of β_k is:

$$\begin{aligned}
&\pi(\beta_k|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{\theta}_{-\beta_k}) \\
\propto &L(\beta_k, \boldsymbol{\theta}_{-\beta_k}, \boldsymbol{R_1}, \boldsymbol{R_2}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2}) \pi(\beta_k, \boldsymbol{\theta}_{-\beta_k}) \\
\propto &\left[\prod_{i=1}^{N}\prod_{j=1}^{n_i} \exp\left[\frac{-\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}^2}{2\sigma^2}\right]\right] \exp\left\{\frac{-(\beta_k - \mu_{\beta_k})^2}{2\sigma_{\beta_k}^2}\right\}.
\end{aligned}$$

Hence kernel of the logarithm of the conditional distribution is

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{-\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}^2}{2\sigma^2}\right] - \frac{(\beta_k - \mu_{\beta_k})^2}{2\sigma_{\beta_k}^2}.$$

Then

$$\frac{d\log\pi(\beta_k|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{\theta}_{-\beta_k})}{d\beta_k} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}x_{ijk}}{\sigma^2} - \frac{\beta_k - \mu_{\beta_k}}{\sigma_{\beta_k}^2},$$

and

$$\frac{d^2\log\pi(\beta_k|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{\theta}_{-\beta_k})}{d\beta_k^2} = -\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{x_{ijk}^2}{\sigma^2} - \frac{1}{\sigma_{\beta_k}^2}. \tag{A.2}$$

Since both terms in (A.2) are negative, we have the full conditional distribution of β_k $(k = 0, ..., K-1)$ is log-concave.

γ:

The full conditional distribution of γ is:

$$\begin{aligned}
&\pi(\gamma|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{\theta}_{-\gamma}) \\
\propto &L(\gamma, \boldsymbol{\theta}_{-\gamma}, \boldsymbol{R_1}, \boldsymbol{R_2}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) f_{\boldsymbol{R_i}}(r_{1i}, r_{2i}|\boldsymbol{\theta_2}) \pi(\gamma, \boldsymbol{\theta}_{-\gamma}) \\
\propto &\left[\prod_{i=1}^{N}\prod_{j=1}^{n_i} \exp\left[\frac{-\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}^2}{2\sigma^2}\right]\right] \exp\left\{\frac{-(\gamma - \mu_{\gamma})^2}{2\sigma_{\gamma}^2}\right\}.
\end{aligned}$$

Hence kernel of the logarithm of the conditional distribution is

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{-\{y_{2ij} - x_{ij}^T\boldsymbol{\beta} - \gamma(y_{1ij} - p_{1ij}) - r_{2i}\}^2}{2\sigma^2}\right] - \frac{(\gamma - \mu_{\gamma})^2}{2\sigma_{\gamma}^2}.$$

Then

$$\frac{d\log\pi(\gamma|\boldsymbol{Y_1},\boldsymbol{Y_2},\boldsymbol{\theta}_{-\gamma})}{d\gamma}$$
$$=\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\{y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma(y_{1ij}-p_{1ij})-r_{2i}\}(y_{1ij}-p_{1ij})}{\sigma^2}-\frac{\gamma-\mu_\gamma}{\sigma_\gamma^2},$$

and

$$\frac{d^2\log\pi(\gamma|\boldsymbol{Y_1},\boldsymbol{Y_2},\boldsymbol{\theta}_{-\gamma})}{d\gamma^2}=-\frac{(y_{1ij}-p_{1ij})^2}{\sigma^2}-\frac{1}{\sigma_\gamma^2}. \tag{A.3}$$

Since both terms in (A.3) are negative, we have the full conditional distribution of γ is log-concave.

R_{1i} $(i=1,...,N)$:

We need to show whether the following distribution is log-concave or not:

$$\prod_{i=1}^{N}\left\{\prod_{j=1}^{n_i}f_{Y_{1ij}}(y_{1ij}|r_{1i})\right\}f_{R_{1i}}(r_{1i}|\phi)$$
$$\propto\prod_{i=1}^{N}\left[\prod_{j=1}^{n_i}\left[\frac{\exp\{(x_{ij}^T\boldsymbol{\alpha}+r_{1i})y_{1ij}\}}{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})}\right]\frac{\sin(\phi\pi)}{2\pi\{\cosh(\phi r_{1i})+\cos(\phi\pi)\}}\right].$$

Kernel of the logarithm of the conditional distribution is

$$\sum_{i=1}^{N}\left[\sum_{j=1}^{n_i}[(x_{ij}^T\boldsymbol{\alpha}+r_{1i})y_{1ij}-\log\{1+\exp(x_{ij}^T\boldsymbol{\alpha}+r_{1i})\}]-\log\{\cosh(\phi r_{1i})+\cos(\phi\pi)\}\right].$$

Then the first derivative of the logarithm of the conditional distribution with respect to R_{1m} $(m=1,...,N)$ is

$$\sum_{j=1}^{n_m}\left\{y_{1mj}-\frac{\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}{1+\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}\right\}-\frac{\phi\sinh(\phi r_{1m})}{\cosh(\phi r_{1m})+\cos(\phi\pi)},$$

and the second derivative is

$$
\begin{aligned}
&-\sum_{j=1}^{n_m}\left[\frac{\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}{1+\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}-\frac{\exp(2x_{mj}^T\boldsymbol{\alpha}+2r_{1m})}{\{1+\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})\}^2}\right]- \\
&\left[\frac{\phi^2\cosh(\phi r_{1m})}{\cosh(\phi r_{1m})+\cos(\phi\pi)}-\frac{\phi^2\sinh^2(\phi r_{1m})}{\{\cosh(\phi r_{1m})+\cos(\phi\pi)\}^2}\right] \\
=&-\sum_{j=1}^{n_m}\frac{\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}{\{1+\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})\}^2}-\phi^2\frac{\cosh^2(\phi r_{1m})+\cosh(\phi r_{1m})\cos(\phi\pi)-\sinh^2(\phi r_{1m})}{\{\cosh(\phi r_{1m})+\cos(\phi\pi)\}^2} \\
=&-\sum_{j=1}^{n_m}\frac{\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})}{\{1+\exp(x_{mj}^T\boldsymbol{\alpha}+r_{1m})\}^2}-\phi^2\frac{1+\cosh(\phi r_{1m})\cos(\phi\pi)}{\{\cosh(\phi r_{1m})+\cos(\phi\pi)\}^2} \qquad \text{(A.4)}
\end{aligned}
$$

The first term in (A.4) is negative. However, the second term is negative when $0<\phi<1/2$ since $\cosh(\phi r_{1m})>0$ and $\cos(\phi\pi)>0$, and may not be negative when $1/2<\phi<1$, i.e. $\cos(\phi\pi)<0$. Hence Metropolis-Hastings algorithm is used for sampling R_{1i} $(i=1,...,N)$ which has infinite range.

R_{2i} $(i=1,...,N)$:

To check the log-concavity of R_{2i} $(i=1,...,N)$, we need to show whether the following distribution is log-concave or not:

$$
\begin{aligned}
&\prod_{i=1}^{N}\left\{\prod_{j=1}^{n_i}f_{Y_{2ij}}(y_{2ij}|r_{2i})\right\}f_{R_{2i}}(r_{2i}|\sigma_r^2) \\
\propto&\prod_{i=1}^{N}\left[\prod_{j=1}^{n_i}\exp\left[\frac{-\{y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma(y_{1ij}-p_{1ij})-r_{2i}\}^2}{2\sigma^2}\right]\exp\left(-\frac{r_{2i}^2}{2\sigma_r^2}\right)\right].
\end{aligned}
$$

Kernel of the logarithm of the above conditional distribution for R_{2i} $(i=1,...,N)$ is given as follows

$$
\sum_{i=1}^{N}\left[\sum_{j=1}^{n_i}\frac{-\{y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma(y_{1ij}-p_{1ij})-r_{2i}\}^2}{2\sigma^2}-\frac{r_{2i}^2}{2\sigma_r^2}\right].
$$

Then the first derivative of the logarithm of the conditional distribution with respect to R_{2m} $(m=1,...,N)$ is

$$
\sum_{j=1}^{n_m}\frac{y_{2mj}-x_{mj}^T\boldsymbol{\beta}-\gamma(y_{1mj}-p_{1mj})-r_{2m}}{\sigma^2}-\frac{r_{2m}}{\sigma_r^2},
$$

and the second derivative is

$$-\sum_{j=1}^{n_m}\frac{1}{\sigma^2}-\frac{1}{\sigma_r^2}. \tag{A.5}$$

Since both terms in (A.5) are negative, we have the full conditional distribution of R_{2i} $(i=1,...,N)$ is log-concave.

The full conditional distribution of ϕ is proportional to

$$\left\{\prod_{i=1}^{N} f_{R_{1i}}(r_{1i}|\phi)\right\}\pi(\phi)\propto\left\{\prod_{i=1}^{N}\frac{\sin(\phi\pi)}{\cosh(\phi r_{1i})+\cos(\phi\pi)}\right\}\cdot 1.$$

The full conditional distribution of ρ_2 is proportional to

$$\left\{\prod_{i=1}^{N} f_{\boldsymbol{R_i}}(r_{1i},r_{2i}|\rho_2)\right\}\pi(\rho_2)\propto\left[\prod_{i=1}^{N}\frac{1}{\sqrt{1-\rho_2^2}}\exp\left\{-\frac{z_1^2-2\rho_2 z_1 z_2+z_2^2}{2(1-\rho_2^2)}\right\}\right]\cdot\frac{1}{2},$$

where $Z_{1i}=\Phi^{-1}(F_\phi(R_{1i}))$ and $Z_{2i}=R_{2i}/\sigma_r$.

The full conditional distribution of σ_r^2 is proportional to

$$\left\{\prod_{i=1}^{N} f_{R_{2i}}(r_{2i}|\sigma_r^2)\right\}\pi(\sigma_r^2)\propto\left\{\prod_{i=1}^{N}(\sigma_r^2)^{-1/2}\exp(-\frac{r_{2i}^2}{2\sigma_r^2})\right\}(\sigma_r^2)^{-a-1}\exp(-\frac{b}{\sigma_r^2}).$$

The full conditional distribution of σ^2 is proportional to

$$\begin{aligned}&L(\sigma^2,\boldsymbol{\theta}_{-\sigma^2},\boldsymbol{R_1},\boldsymbol{R_2}|\boldsymbol{Y_1},\boldsymbol{Y_2})f_{\boldsymbol{R_i}}(r_{1i},r_{2i}|\boldsymbol{\theta_2})\pi(\sigma^2,\boldsymbol{\theta}_{-\sigma^2})\\ \propto&\prod_{i=1}^{N}\left[\prod_{j=1}^{n_i}(\sigma^2)^{-1}\exp\left[-\frac{\{y_{2ij}-x_{ij}^T\boldsymbol{\beta}-\gamma(y_{1ij}-p_{1ij})-r_{2i}\}^2}{2\sigma^2}\right]\right](\sigma^2)^{-c-1}\exp(-\frac{d}{\sigma^2}).\end{aligned}$$

We have evaluated the above four conditional distributions, however, these distributions are not log-concave.

APPENDIX B

WINBUGS CODES AND SAS CODE FOR MODEL 3

The Winbugs Codes for the parametric Bayesian method and the semiparametric Bayesian method as well as the SAS code for MLE of Model 3 are given in this Appendix.

Winbugs code for parametric Bayesian method of Model 3:

```
model
{
    zmu[1]<-0
    zmu[2]<-0
    cova[1,1]<-1
    cova[2,2]<-1
    cova[1,2]<-rou
    cova[2,1]<-cova[1,2]
    prec[1:2,1:2]<-inverse(cova[,])

    for(i in 1:M)
    {
    z[i,1:2]~ dmnorm(zmu[1:2],prec[1:2,1:2])
    v[i]<-phi(z[i,1])

    R1[i]<-1/phi*log(sin(phi*3.1416*v[i])/sin(phi*3.1416*(1-v[i])))
    R2[i]<-z[i,2]
    }
```

```
for (j in 1:N)
{
y[j,6]∼dbern(p[j])
logit(p[j])<-alpha0+alpha1∗y[j,4]+R1[y[j,2]]
logit(p1[j])<-alpha0∗phi+alpha1∗phi∗y[j,4]

y[j,5]∼dnorm(u[j],tau)
u[j]<-beta0+beta1∗y[j,4]+gamma∗(y[j,6]-p1[j])+sigmar∗R2[y[j,2]]
}

rou∼dunif(-0.999,0.999)
phi∼dunif(0.001,0.999)

taur∼dgamma(1,1)
sigmar<-sqrt(1/taur)
tau∼dgamma(1,1)
sigma<-sqrt(1/tau)

alpha0∼dnorm(-4.6,0.1)
alpha1∼dnorm(0,0.1)
beta0∼dnorm(0.9,0.1)
beta1∼dnorm(0,0.1)
gamma∼dnorm(0,0.1)

for (j in 1:N)
{
logly[j]<-log(pow(p[j],y[j,6])∗pow(1-p[j],1-y[j,6])∗1/sqrt(2∗3.1416∗sigma∗sigma)
∗exp(-(y[j,5]-u[j])∗(y[j,5]-u[j])/(2∗sigma∗sigma)))

logsum[j]<-sum(logly[y[j,7]:y[j,8]])
CPOclust[j]<-1/(exp(logsum[j]))
}
```

```
}

list
(
    M=94, N=1028, A=10, y = structure(.Data = c(60,1,7,0,0.903,0,1,7,
    . . . . . .
    156,94,12,3,0.829,0,1017,1028),
    .Dim = c(1028,8))
)
```

Winbugs code for semiparametric Bayesian method of Model 3:

```
model
{
    zmu[1]<-0
    zmu[2]<-0
    cova[1,1]<-1
    cova[2,2]<-1
    cova[1,2]<-rou
    cova[2,1]<-cova[1,2]
    prec[1:2,1:2]<-inverse(cova[,])

    for(i in 1:M)
    {
    z[i,1:2]~ dmnorm(zmu[1:2],prec[1:2,1:2])
    v[i]<-phi(z[i,1])

    R1[i]<-1/phi*log(sin(phi*3.1416*v[i])/sin(phi*3.1416*(1-v[i])))
    R2[i]<-z[i,2]
    }
```

```
for (j in 1:N)
{
y[j,6]∼dbern(p[j])
logit(p[j])<-alpha0+alpha1∗y[j,4]+R1[y[j,2]]
logit(p1[j])<-alpha0∗phi+alpha1∗phi∗y[j,4]

y[j,5]∼dnorm(u[j],tau)
u[j]<-beta0+beta1∗y[j,4]+gamma∗(y[j,6]-p1[j])+sigmar∗R2[y[j,2]]
tau[j]<-1/(sigma[j]*sigma[j])
sigma[j]<-sigmastar[k[j]]

k[j]∼dcat(probi[])
}

C∼dgamma(1,0.1)

prob[1]<-L[1]

for (i in 2:A)
{
prob[i]<-L[i]∗(1-L[i-1])∗prob[i-1]/L[i-1]
}

prob.sum<-sum(prob[])

for (h in 1:A)
{
L[h]∼dbeta(1,C)
probi[h]<-prob[h]/prob.sum

sigmastar[h]<-1/sqrt(taubase[h])
taubase[h]∼dgamma(1,1)
```

```
    }

    rou~dunif(-0.999,0.999)
    phi~dunif(0.001,0.999)

    taur~dgamma(1,1)
    sigmar<-sqrt(1/taur)

    alpha0~dnorm(-4.6,0.1)
    alpha1~dnorm(0,0.1)
    beta0~dnorm(0.9,0.1)
    beta1~dnorm(0,0.1)
    gamma~dnorm(0,0.1)

    for (j in 1:N)
    {
    logly[j]<-log(pow(p[j],y[j,6])*pow(1-p[j],1-y[j,6])*1/sqrt(2*3.1416*sigma[j]*sigma[j])
    *exp(-(y[j,5]-u[j])*(y[j,5]-u[j])/(2*sigma[j]*sigma[j])))

    logsum[j]<-sum(logly[y[j,7]:y[j,8]])
    CPOclust[j]<-1/(exp(logsum[j]))
    }

}

list
(
    M=94, N=1028, A=10, y = structure(.Data = c(60,1,7,0,0.903,0,1,7,
    . . . . . .
    156,94,12,3,0.829,0,1017,1028),
    .Dim = c(1028,8))
)
```

SAS code for MLE of Model 3:

```
data EGdata;
set 'C:\EGdata';
run;

data dataz(keep=litterid dose malformation weight zzz);
set EGdata;
zzz=1;
run;

ods listing;

proc nlmixed
data=dataz method=ISAMP qpoints=200 maxiter=100 tech=NRRIDG NOAD;

pi=constant('pi');

uni=probnorm(z1);
phi=1.0/sqrt(1+3/pi/pi*stdrebri*stdrebri);
Bbin=1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));

xbbin=Bbin+a0+a1*dose;

p=exp(xbbin)/(1+exp(xbbin));
p1=exp(phi*(xbbin-Bbin))/(1+exp(phi*(xbbin-Bbin)));

llik1=malformation*log(p)+(1-malformation)*log(1-p);

xbnor=stdrenor*z2+b0+b1*dose+g1*(malformation-p1);

llik2=((-0.5)*log(2*pi*signor*signor)-((weight- xbnor)**2)/(2*signor*signor));
```

```
llik=llik1+llik2;

model zzz~general(llik);

random z1 z2 ~normal([0,0],[1,rhobrinor,1]) subject=litterid;

ESTIMATE 'a0mar' phi*a0;
ESTIMATE 'a1mar' phi*a1;
ESTIMATE 'phi' phi;
ESTIMATE 'rhobri' 1-phi;
ESTIMATE 'rhonor' (stdrenor*stdrenor)/ (signor*signor+stdrenor*stdrenor);

ods output ParameterEstimates=pars;
ods output AdditionalEstimates=margest;

run;
```

REFERENCES

[1] D.R. Cox and N. Wermuth. Response models for mixed binary and quantitative variables. *Biometrika*, 79(3):441–461, 1992. 1.1, 1.2

[2] M.D. Sammel, L.M. Ryan, and J.M. Legler. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 59(3):667–678, 1997. 1.1

[3] D.B. Dunson. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 62(2):355–366, 2000. 1.1, 1.2, 1.3

[4] A.R. de Leon and K.C. Carriere. General mixed-data model: Extension of general location and grouped continuous models. *The Canadian Journal of Statistics*, 35(4):533–548, 2007. 1.1, 7

[5] C.J. Price, C.A. Kimmel, R.W. Tyl, and M.C. Marr. The developmental toxicity of ethylene glycol in rats and mice. *Toxicology and Applied Pharmacology*, 81(1):113–127, 1985. 1.1, 4.1

[6] G.M. Fitzmaurice and N.M. Laird. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association*, 90(431):845–852, 1995. 1.1, 1.2, 6.1

[7] P. Catalano and L.M. Ryan. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, 87(419):651–658, 1992. 1.2

[8] M.M. Regan and P.J. Catalano. Likelihood models for clustered binary and continuous outcomes: Applications to developmental toxicology. *Biometrics*, 55(3):760–768, 1999. 1.2

[9] H. Geys, G. Molenberghs, and L.M. Ryan. Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, 94(447):734–745, 1999. 1.2, 6.2

[10] R. Gueorguieva and Alan. Agresti. A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96(455):1102–1112, 2001. 1.2

[11] Z. Wang and T. A. Louis. Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika*, 90(4):765–775, 2003. 1.3, 2.2.2

[12] J. Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994. 1.3, 5.3

[13] M.D. Escobar and M. West. Computing nonparametric hierarchical models. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, pages 1–22. Springer, New York, 1998. 1.3, 5.1, 5.2

[14] A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, New York, second edition, 2002. 2.1

[15] R. B. Nelsen. *An Introduction to Copulas*. Springer, New York, 1999. 2.4.1

[16] A. Sklar. Fonctions de rpartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universit de Paris*, 8:229–231, 1959. 2.4.1

[17] P. Hougaard. *Analysis of Multivariate Survival Data*. Springer, New York, 2000. 2.4.2

[18] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, second edition, 2000. 3

[19] P. Congdon. *Bayesian Statistical Modelling*. John Wiley and Sons, Chichester, second edition, 2006. 3, 5.3

[20] D. Gamerman and H.F. Lopes. *Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, second edition, 2006. 3.1.1

[21] A. Srivastava. Computational methods in statistics. Class notes, Department of Statistics, Florida State University. 3.1.1

[22] W.R. Gilks and P. Wild. Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, 41(2):337–348, 1992. 3.1.2

[23] R.M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003. 3.1.2, 3.1.2, 4.3

[24] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21(6):1087–1092, 1953. 3.1.2, 3.1.2, 4.3

[25] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. 3.1.2, 3.1.2, 4.3

[26] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984. 3.1.2, 3.1.2

[27] P. Muller. Metropolis based posterior integration schemes. Technical report, Statistics Department, Purdue University. 3.1.2, 3.1.2

[28] W.R. Gilks. Derivative-free adaptive rejection sampling for gibbs sampling. In J. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 641–649. Oxford University Press, Oxford, 1992. 3.1.2, 4.3

[29] A.E. Gelfand, D.K. Dey, and H. Chang. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In J. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 641–649. Oxford University Press, Oxford, 1992. 3.5

[30] J.G. Ibrahim, M.H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, New York, 2001. 3.5

[31] D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337, 2000. 4.3, 5.3

[32] T. Ferguson. A bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1(2):209–230, 1973. 5.1

[33] M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995. 5.2

[34] W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979. 6.1

[35] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. 6.2

[36] G.E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978. 6.2

[37] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, second edition, 1998. 7

BIOGRAPHICAL SKETCH

Lanjia Lin

Lanjia Lin was born on July 10th, 1981, Wuhan, China. In the fall of 1999, she attended Wuhan University in China and completed her Bachelor's degree in Computer Science in the summer of 2003. She studied Mathematics at University of South Carolina, obtaining a Masters of Science degree in the summer of 2005. She enrolled in the doctoral program in Statistics at Florida State University in the fall of 2005. She defended her dissertation in the spring of 2009.

Lanjia's current research interests include modeling of clustered data with mixed responses, categorical data analysis, semiparametric Bayesian analysis, developmental toxicity study and clinical trials.