University of Southampton

Faculty of Mathematical Studies

Mathematics

# Modelling Ordinal Categorical Data : A Gibbs Sampler Approach

by

Wan-Kai Pang

Thesis submitted for the degree of Doctor of Philosophy

March, 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES

MATHEMATICS

Doctor of Philosophy

# Modelling Ordinal Categorical Data : A Gibbs Sampler Approach

by Wan-Kai Pang

This thesis presents a study of statistical models for ordered categorical data. The generalized linear model plays an essential role in this approach. A Gibbs sampler method is used to estimate model parameters for a Bayesian formulation of a random effects generalized linear model. The adaptive rejection sampling (ARS) method introduced by Gilks and Wild (1992) is used in the Gibbs sampling scheme. Good results are obtained in simulations and we applied this model to analyze data concerning telephone connection quality supplied by British Telecom (BT). The concept of latent residuals introduced by Albert and Chib (1995) is used for diagnostic checking.

A random effects cumulative logit model is employed to analyze longitudinal ordinal responses and a Bayesian approach to the cumulative logit model is considered. The adaptive rejection sampling (ARS) technique is again used to estimate model parameters. Simulation results as well as results from a real application are presented. A new cumulative logit model is developed to cater for a particular set of ordinal categorical data. The main reason is that in the telephone connection quality experiment, each subject has his/her personal scale in mind. At the same time, the underlying stochastic ordering structure needs to be maintained for the model. This model is used to model the telephone connection quality data. A continuation-ratio model and cumulative probit model with serial correlation are also considered.

(i) Material from this thesis also appeared in the paper entitled– "Modelling Binary Data: A Gibbs Sampling Approach" by Wan-Kai Pang in *Journal of Statistical Computation and Simulation.* (1999), **64**, pp73-85.

(ii) Material from this thesis also appeared in the *Proceedings in Computational Statistics*; *Short Communications and Posters*, 1998, pp91-92; edited by R. Payne and P. Lane.

(iii) The FORTRAN codes written in this thesis are available from the author on request.

# Contents

# List of Tables

# List of Figures

# ACKNOWLEDGEMENTS

*Sincere thanks to my supervisor, Dr. Jonathan J. Forster, whose intelligence, intuition and constant encouragement made this work possible. Also I wish to express my sincere thank to Dr. Gilks of MRC-Biostatistics unit, Cambridge, for sending me the Fortran code for the ARS method and I also thank British Telecom for giving me the permission to use their data. I would like to thank The Hong Kong Polytechnic University for partially supporting my studies and to thank all of my friends and colleagues who once helped me in finishing my PhD study.*

*I also thank Professor Sue Lewis and Professor Phil Prescott for their helpful comments and suggestions.*

*Finally, my sincere thank goes to my wife for her support on my studies for the past six years.*

# Chapter 1

# Introduction

## 1.1 An Overview of Existing Literature on Modelling Repeated Ordinal Data

In many statistical studies, responses are measured repeatedly for each subject under investigation. Correlation is common amongst the repeated measures series. In this thesis, I consider situations where the responses are recorded on an ordinal scale. An ordinal scale response variable is also referred to as an ordered categorical response variable. For example in a clinical trial, the responses of a patient to a sequence of trials on certain drug may be measured on a four-point scale: excellent, good, fair and poor.

### 1.1.1 Ordinal Responses

An approach to analyzing ordinal responses which does not involve assigning an arbitrary score to each category is to view the ordinal response $y_{ij}$ of subject $i$ on occasion $j$ as the ordinal categorical manifestation of an underlying continuous random variable. Suppose there exists a latent continuous variable $\lambda_{ij}$, then the observation $y_{ij}$

is observed in the following manner:

$$y_{ij} = h \qquad \text{if} \qquad \alpha_{h-1} < \lambda_{ij} < \alpha_h \qquad (h = 1, 2, \ldots, k), \tag{1.1}$$

where $\alpha_h$, $h = 1, 2, \ldots, k$ are called the cutpoint parameters. There are $k - 1$ cutpoints. Now let $\lambda_{ij}$ take the linear form $\lambda_{ij} = \underline{\beta}'\underline{x}_i + \epsilon_{ij}$ where $\epsilon_{ij}$ has a specific distribution $G$ and $E(\epsilon_{ij}) = 0$. $\underline{x}_i$ is the appropriate row of the general design matrix $X$. The parameters to be estimated are contained in the vector $\underline{\beta}$. Therefore in this setting we have

$$P(y_{ij} \leq h) = G(\lambda_{ij} - \underline{\beta}'\underline{x}_i) \tag{1.2}$$

The $\alpha_h s$ are necessarily ordered with $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_k = \infty$. Different choices of $G^{-1}$ will yield different models for cumulative probabilities. For instance if $\lambda_{ij} = \underline{\beta}'\underline{x}_i + \epsilon_{ij}$ with $\epsilon$ having a logistic distribution, then $G^{-1}$ is the logit transform. This and related models have been extensively discussed by McCullagh (1980) and more recently by Agresti (1996). The model in equation (1.2) with a logit link is called the fixed effect cumulative logit model. The random effects cumulative logit is simply of the form

$$\lambda_{ij} = \underline{\beta}'\underline{x}_i + \underline{b}_i'\underline{z}_i + \epsilon_{ij}, \tag{1.3}$$

where $\underline{z}_i$ is the random effects component for each subject $i$ and it is a subset of $\underline{x}_i$. Now let $\gamma_{ij} = P(y_{ij} \leq h)$, then the following list shows different models for cumulative probabilities with different link functions:

(i) Cumulative Logit Model: $G^{-1}(\gamma_{ij}) = \log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right)$

(ii) Probit Model: $G^{-1}(\gamma_{ij}) = \Phi^{-1}(\gamma_{ij})$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

(iii) Complementary log-log Model: $G_C^{-1} = \log\{-\log(1-\gamma_{ij})\}$, where $G$ is assumed to be the standard extreme minimum value distribution.

(iv) Log-log Model: $G_L^{-1} = \log\{-\log(\gamma_{ij})\}$, where $G$ is assumed to be the standard extreme maximum value distribution and the relation $G_C^{-1} = -G_L^{-1}(1-\gamma_{ij})$ holds.

There are other types of cumulative probability models for ordinal responses. For example, the *continuation-ratio*, probability,

$$\gamma_{ij} = P(y_{ij} = h \mid y_{ij} \geq h) = \frac{\pi_{ijh}}{\pi_{ijh} + \cdots + \pi_{ijk}}, \quad h = 1, 2, \ldots, k. \tag{1.4}$$

A logit link for $\gamma_{ij}$ in equation 1.4 is referred to as the *continuation-ratio* logit model:

$$\text{logit}\left(\frac{\pi_{ijh}}{1 - \gamma_{ijh}}\right) = \alpha_h - \underline{\beta}'\underline{x}_i, \tag{1.5}$$

where $\pi_{ijh} = \Pr(y_{ij} = h)$ and $\gamma_{ijh} = \pi_{ij1} + \pi_{ij2} + \cdots + \pi_{ijh}$.

Similarly there are the continuation-ratio complementary log-log model and continuation-ratio probit model when the complementary log-log and probit link functions are used respectively.

Another ordinal probability model is the *adjacent-category* probability model

$$\xi_{ijh} = \frac{\pi_{ijh}}{\pi_{ijh} + \pi_{ij(h+1)}} \tag{1.6}$$

This is the probability of an observation being in category $h$ given that it is in category $h$ or $h+1$. Again taking different link functions would result in different models. Models for analyzing ordinal data in this thesis are basically cumulative logit models. Cumulative logit models have the advantage of being easy to interpret. One can calculate the log-odds ratio concerning a covariate using the regression coefficient vector $\underline{\beta}$. One example of interpretation of the log-odds ratio is from Francom, Chuang and Landis (1989). They showed an example of repeated ordered categorical response data in tabular form. The table gave results of a randomized, double-blind clinical trial comparing an active hypnotic drug with a placebo in patients with insomnia. The outcome variable was patient response to the question 'How quickly did you fall asleep after going to bed ? ' using categories ( <20, 20-30, 30-60 >60 ) minutes. Patients responded at the start and conclusion of a two-week treatment period. The repeated measurement makes the response bivariate, measured at levels: initial and follow-up. After fitting an interaction model to the data, they estimated that for the placebo treatment, the odds that time to falling asleep below any fixed level is approximately equal to $\exp(1.05) = 2.9$. This indicates that the odds of time to fall asleep at the follow-up occasion is almost three times higher than that at the initial trial. For the active treatment, the effect is $\exp(1.05 + 0.65) = 5.5$. More details about the interpretation of model parameters are provided by McCullagh (1980).

The cumulative logit has the advantage of a certain invariance to response category choice. If a cumulative logit model holds for an underlying continuous responses, it also holds for any categorical measurement of the responses, with the same values for the effect parameters. For sample data, if the model fits well for a fixed set of response categories, it also tends to fit well when we combine sets of adjacent responses. When there is an arbitrary rather than a fixed choice of response categories, interpretation of parameters may also be more natural for cumulative logit models. When there is a fixed set of responses, the adjacent-category logit is sometimes more useful, since it permits contrasts between pairs of response categories.

In chapter 6, a probit link function is proposed for a model which includes an autoregressive process in its linear functional form. A probit link will facilitate the estimation procedure. However in many situations, a logit or probit link will both fit well. The only real difference is the measurement scale of the parameters.

### 1.1.2 Estimation for Ordinal Data

Now suppose in an experiment, each subject may be observed at $d$ occasions, and let $(1, 2, \ldots, r)$ denote the $r$ possible response categories at each occasion. The data can be described by a contingency table with $r^d$ cells, containing counts of possible multivariate response profiles, Let

$$\underline{\pi}_{\mathbf{j}} \qquad \text{with } \mathbf{j} = (j_1, \ldots, j_d) \tag{1.7}$$

denote the probability that a random selected subject makes response $j_g$ at occasion $g$, $1 \leq j_g \leq r$, $g = 1, \ldots, d$. Ashford (1959), Cox (1970) and McCullagh (1980) considered models for categorical response variables in general. Haber (1985) gave an iterative Newton-Raphson routine of obtaining maximum likelihood estimates (MLEs) of parameters in the models of the form

$$\mathbf{A}\log\mathbf{B}\underline{\pi} = \underline{x}'\underline{\beta} \tag{1.8}$$

where $\underline{\pi}$ denotes the cell probabilities of a multinomial sample over the $r^d$ possible response profiles, with independent samples at each of the $s$ level of $\underline{x}$ where $\underline{x}$ is the covariate vector. For adjacent-category logits, $\mathbf{B}$ contains '0' and '1' elements such that $\mathbf{B}\underline{\pi}$ produces the $rd$ marginal probabilities for each level of $\underline{x}$ and each row of the matrix $\mathbf{A}$ contains '0' elements except for a single '1' and '−1' positioned to form a particular logit. For cumulative logits, $\mathbf{B}$, produces the $2(r-1)d$ cumulative probabilities and their complements. Haber (1985) uses Aitchison and Silvey's (1958) method for maximizing a likelihood subject to constraints. However Haber's routines are impractical when the table has a large number of cells because of the problem of large matrix inversion.

Koch *et al.* (1977) use the model in equation (1.8) for repeated measure categorical data. A weighted least square (WLS) approach is suggested for parameter estimation. Their estimation procedure is outlined as follows:

Let $\underline{p}$ be the sample proportion estimate of $\underline{\pi}$ and $\text{Var}(\underline{p}) = V$. When $\underline{x}$ has $s$ levels, $V$ is an $s$-block diagonal matrix with separate multinomial covariance for each block. The WLS estimate of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = (\underline{x}'S^{-1}\underline{x})^{-1}\underline{x}'S^{-1}(\mathbf{A}\mathbf{B}\underline{p}) \tag{1.9}$$

where $S = \mathbf{A}D^{-1}\mathbf{B}V\mathbf{B}'D^{-1}\mathbf{A}'$ is the approximate covariance matrix for the model by using the delta method. The asymptotic covariance of $\hat{\underline{\beta}}$ is given by the matrix $(\underline{x}'S^{-1}\underline{x})^{-1}$. The quadratic form

$$(\mathbf{A}\text{log}\mathbf{B}\underline{p} - \underline{x}'\underline{\beta})'S^{-1}(\mathbf{A}\text{log}\mathbf{B}\underline{p} - \underline{x}'\underline{\beta}) \tag{1.10}$$

is used for testing goodness of fit. Wald statistics are used for hypothesis testing. A disadvantage of WLS is its inefficiency in handling continuous covariates. Missing data or time-dependent covariates often make the WLS procedure complicated. Stram, Wei and Ware (1988) proposed a semi-parametric approach for fitting a cumulative logit model to repeated measures data. They assume no dependence structure among the repeated observations. Their approach is mainly focused on estimation of covariate effects rather than occasion effects. This yields estimates of cutpoint parameters $\alpha_k$

for the various occasions. One could also use a semi-parametric approach to obtain estimated covariances for the cutpoint estimates. In principle one can extend this semi-parametric methodology for alternative links, such as adjacent-category logits.

Another semi-parametric approach can be developed for repeated binary responses. This is an extension of the methodology proposed by Liang and Zeger (1986). The model parameters are estimated as if the repeated observations were independent. The parameter estimates are consistent and asymptotically normal, but the inverse of the estimated information matrix is not consistent for the true asymptotic covariance matrix. The ML, WLS and semi-parametric approaches each have certain advantages. The ML and WLS approaches have the elegance of simultaneously describing occasion and covariate effects. The semi-parametric approaches may be less efficient than ML or WLS approaches in estimating effects if multinomial model holds. The semi-parametric approaches make it simpler to allow for time-dependent covariates and for missing data as compared to ML and WLS approaches. For cumulative logit models, Landis *et al.* (1987) incorporated sampling weights and design effects into test statistics using Taylor-series approximations to obtain weighted proportions and their covariance matrix. This work is from a classical frequentist approach. One may also refer to the books by Agresti (1996) and Fienberg (1994) for more comprehensive work on models for categorical data. Jansen (1990), McCullagh (1977, 1978) and Stram, Wei and Ware (1988) have published articles focusing on models for ordinal data in which the responses may be clustered. For example, in studies involving related individuals, repeated or multiple measurement on each individual are recorded. In such studies the ordered responses of the subjects within the clusters can be positively correlated. One way to explain this correlation is to postulate that the linear predictors for each clustered response share cluster-specific effects. Agresti and Lang (1993) presented models for repeated ordered categorical responses with subject-specific effects. Their model fitting process uses an improved Newton-Raphson algorithm for fitting generalized loglinear models by maximum likelihood estimation subject to constraints.

Albert and Chib (1993) used exact Bayesian methods for modelling binary and polychotomous response data. Binary data is simply a special kind of categorical data in which there are only two categories, namely, success and failure. The work of Albert and Chib (1993) is based on the concept of a latent variable. This employed the Gibbs sampler for parameter estimation.

The work of this thesis is largely based on these ideas and a Bayesian approach to statistical modelling is adopted. This thesis gives details of unified Bayesian approach for modelling ordinal data with random effects. We use an efficient Gibbs sampler scheme to estimate the model parameters. This is especially appropriate for estimating the random effects components of the model. Estimating the random effects component can be a difficult task in a classical approach (see Pan and Thompson, 1998) as well as for the Gibbs sampler when using ordinary rejection sampling method (see Zeger and Karim, 1991). The sampling method used in this thesis is the adaptive rejection sampling (ARS) method developed by Gilks and Wild (1992). We develop a model to cater for ordinal scale data. The new model is useful for some experiments in which the subjects under observation have their own scale of measurements in mind. Finally, a model is proposed to cater for longitudinal ordinal responses where time dependence may occur of one trial on another for each subject $i$. The overall results for the models are good in simulation studies. The models are also applied to real data from telecommunications experiments. For diagnostic checking of the models of empirical data, we examine the latent residuals. The concept of latent residuals is introduced by Albert and Chib (1995). Details of latent residuals can be found in section (2.4.2) of chapter 2. In the following sections, some background on generalized linear models, Bayesian statistical modelling and Markov chain Monte Carlo are briefly introduced.

### 1.1.3 Generalized Linear Model with Random Effects

In recent years, generalized linear models (McCullagh and Nelder, 1989; Nelder and Wederburn, 1972 ) have gained wide popularity in various fields of statistical research as well as in practical applications. The models have unified regression methodology for a wide variety of discrete, continuous, and censored responses that can be assumed to be independent. However in many situations this is unlikely to be the case. For example, in the analysis of repeated measures data, repeated responses from a subject are likely to be correlated. Correlation among responses in a sample often appears in many real life situations. Therefore, dependence must be considered in order to assess the relationship of the response Y with the explanatory variables X and a model which includes random effects terms may be more appropriate than the ordinary generalized linear model. The Gaussian linear model with random effects has the general form

$$y_{ij} = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i + \varepsilon_{ij} \qquad i = 1, ..., n, \;\; j = 1, ..., k, \tag{1.11}$$

where $y_{ij}$ is the response for the $j$th observation in cluster i, $\underline{x}_{ij}$ is a p × 1 vector of covariates associated with that response; $\underline{\beta}$ is the vector of regression coefficients; $\underline{z}_{ij}$ is a q × 1 subset of $\underline{x}_{ij}$ with random coefficients; $\underline{b}_i$ is a q × 1 vector of random effects assumed to follow a Gaussian distribution with mean 0 and unknown variance $D$; and $\varepsilon_{ij}$ is an independent Gaussian error with mean 0 and variance $\sigma^2_\varepsilon$. For a generalized linear model with random effects, conditional on a random component $\underline{b}_i$, $y_{ij}$ follows an exponential family distribution, i.e,

$$f(y_{ij} \mid \underline{b}_i) = \exp\{[y_{ij}\theta_{ij} - a(\theta_{ij}) + c(y_{ij})]/\phi\}. \tag{1.12}$$

The conditional moments $\mu_{ij} = E(y_{ij} \mid \underline{b}_i) = a'(\theta_{ij})$ and $V_{ij} = Var(y_{ij} \mid \underline{b}_i) = a''(\theta_{ij})\phi$ satisfy

$$h(\mu_{ij}) = \eta_{ij} = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b_i} \tag{1.13}$$

and

$$V_{ij} = g(\mu_{ij})\phi \tag{1.14}$$

where $h$ and $g$ are known link and variance functions respectively. If the distribution is normal, the canonical link function is identity. The link refers to a linear combination of $\beta's$ and $b's$ such that the linear combination is equal to some function of the expected value $\mu_{ij}$ of $y_{ij}$, i.e., $h(\mu_{ij}) = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$, where $h$ is a monotone, differentiable function. For the canonical link $h(\mu_{ij}) = \theta_{ij}$. If the distribution is binomial, then the canonical link function is logit. One may refer to the book by McCullagh and Nelder (1989) for more details.

Generalized linear models with random effects have been studied by others in the past twenty years. Examples include Laird and Ware (1982) and Lindstrom and Bates (1988). Both these articles use a maximum likelihood approach to estimate model parameters. Numerical maximization methods such as Newton-Raphson are used. Several authors have investigated the extension of random effects models to the generalized linear model family. Williams (1982) studied the Beta-binomial model and Breslow (1984) studied Poisson-gamma models. Anderson and Aitkin (1985) use EM and Newton-Raphson algorithm to estimate parameters in logistic regression models with a Gaussian random intercept. Gilmour, Anderson and Rae (1985) discussed probit-Gaussian models.

Zeger, Liang and Albert (1988) presented a paper on models for longitudinal data using a generalized estimating equation approach. They considered the generalized linear models with random effects defined in equation 1.13. To model the marginal expectation $\mu_{ij}$, they assume $h^*(\mu_{ij}) = \underline{x}'_{ij}\underline{\beta}^*$ and $Var(y_{ij}) = g^*(\mu_{ij})\phi$. Let $\mu_i = \mathrm{E}(y_i) = \underline{x}'_i\underline{\beta} = \{h^{*-1}(\underline{x}'_{i1}\underline{\beta}^*), \ldots, h^{*-1}(\underline{x}'_{in_i}\underline{\beta}^*)\}'$ and $A_i = \mathrm{diag}\{g^*(\mu_{i1}), \ldots, g^*(\mu_{in_i})\}$. For independent observations, $\mathrm{Cov}(y_i) = A_i\phi$. However we expect correlation among repeated observations for a subject. Let $R_i(\alpha)$ be a "working" correlation matrix. $\underline{\beta}^*$ can be estimated by solving the "generalized estimating equation " (GEE)

$$U(\beta^*) = \sum_{i=1}^{n} \frac{\partial \mu_i'}{\partial \underline{\beta}^*} V_i^{-1}(\alpha)(y_i - \mu_i) = 0 \qquad (1.15)$$

where $V_i(\alpha) = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$. Liang and Zeger (1986) show that $\hat{\underline{\beta}}^*$ is consistent and asymptotically ($n \to \infty$) Gaussian given only correct specification of the mean and the usual regularity conditions. The GEE method can be used in the cumulative logit models for parameter estimation.

Crouchley (1995) presents a random effects models for multivariate and grouped univariate ordered categorical data. He defines a random effects ordered response model for individual $i$'s $j$th response by means of an underlying latent response variable

$$y_{ij} = \beta_0 + \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i + e_{ij} \tag{1.16}$$

where $-\infty < y_{ij} < \infty$, $\beta_0$ is a constant, $\underline{\beta}$ is a vector of unknown parameters, $\underline{x}_{ij}$ is a vector of regressors, $\underline{b}_i$ is a vector of individual-specific random effects, $\underline{z}_{ij}$ is a known matrix, and $e_{ij}$ is a stochastic disturbance term. The term $\underline{b}_i$ and $e_{ij}$ are assumed to be independent.

Unfortunately $y_{ij}$ is not observed, but we do know to which of the $k$-categories it belongs. It is assumed that $y_{ij}$ belongs to the $k$th category if $\alpha_{k-1} < y_{ij} < \alpha_k$ where $\alpha_{k-1}$ and $\alpha_k$ denote the lower and upper boundaries of the $k$ category and also $\alpha_0 = -\infty$, and $\alpha_{k-1} < \alpha_k$ for all $k$. If $G(\cdot)$ is the cumulative distribution function for $e_{ij}$ and if we let $y_{ijk} = 1$ when individual $i$'s $j$th response is in the $k$th category and 0 otherwise, then, conditional on the regressors and the random effects

$$\begin{aligned} P(y_{ijk} = 1 \mid e_{ij}) &= P(\alpha_{k-1} < y_{ij} < \alpha_k \mid e_{ij}) \\ &= P(y_{ij} < \alpha_k \mid e_{ij}) - P(y_{ij} < \alpha_{k-1} \mid e_{ij}) \\ &= G(-\mu_{ij,k} - e_{ij}) - G(-\mu_{ij,k-1} - e_{ij}) \end{aligned} \tag{1.17}$$

where $\mu_{ij,k} = \mu_{ij} - \alpha_k$ and $\mu_{ij} = \beta_0 + \underline{x}'_{ij}\underline{\beta}$. Since $y_{ij}$ is not observed, its location has to be fixed. This is done by setting the constant $\beta_0 = 0$ in $\mu_{ij}$. Different distributions for $e_{ij}$ give different forms for $G(-\mu_{ij,k} - e_{ij})$. If $e_{ij}$ is normal, then we have the ordered probit model; if the $e_{ij}$ is logistic, then we get the ordered logit; if $e_{ij}$ is extreme value distributed, then we get the ordered complementary log-log link model.

Crouchley (1995) presents a random effects ordered response model for the complementary log-log link; that is $G(z) = 1 - \exp\{-\exp(z)\}$. This link function has the advantage that it gives a closed-form expression for the model unconditional on the random effect. Crouchley (1995) assumes that the distribution for the random effects belongs to the Hougaard family (Hougaard, 1986). The Hougaard family is characterized by the Laplace transform, from which many properties such as infinite divisibility and unimodality can be obtained. The density function of the Hougaard family is skewed to the right and in general it can only be presented as an infinite series. Estimation of parameters such as the cut-points $\alpha_k$, and the random effects variance can be done by maximizing the sample log-likelihood function. A quasi-Newton algorithm is used for numerical approximations to the derivatives.

Booth and Hobert (1999) proposed two new implementations of the EM algorithm for maximum likelihood fitting of a generalized linear mixed model. One approach involves generating random samples from the exact conditional distribution of the random effects by rejection sampling. The second method uses a multinomial $t$ importance sampling approximation. Both methods use random (*i. i. d.*) sampling to construct Monte Carlo approximation at the E-step. Monte Carlo approximation using random samples allow the Monte Carlo error at each iteration to be assessed by using standard central limit theory combined with Taylor series methods. A rule for automatically increasing the Monte Carlo sample size after iterations is suggested. The rule for automatically adjusting the size of sample $m$ (say) at the E-step is to construct an approximate $100(1-\alpha)\%$ confidence ellipsoid for the parameter vector $\psi$ (say) at the $(r+1)$th iteration by using the normal approximation derived at section 5 of their article. If the previous value $\psi^{(r)}$ lies in that region, then the EM step was swamped by Monte Carlo error, and $m$ should be increased, e.g. $m \leftarrow m + \frac{m}{k}$, where $k$ is a positive constant. Booth and Hebert (1999) claimed that their method has been successful using with $\alpha = 0.25$ and $k \in \{3, 4, 5\}$. Empirical results in Booth and Hobert's (1999) article show that the methods proposed can be considerably more efficient than those based on Markov chain Monte Carlo algorithms. However the method may break down when the intractable integrals in the likelihood function are of high dimension.

Pan and Thompson (1998) suggested a Quasi-Monte Carlo (QMC) EM algorithm for maximum likelihood estimates in generalized linear mixed models. Inferences for generalized linear mixed models (GLMMs) are often hampered by the intractable integrated likelihood function. The method of Pan and Thompson (1998) uses numerical integration based on Quasi-Monte Carlo method to approximate the integral in the EM algorithm. For generalized linear mixed models, the likelihood function of the regression coefficient vector $\underline{\beta}$ and the variance of the random effects component, $D$, is of the form

$$L(\underline{\beta}, D) = \int f(y \mid \underline{b}_i, \underline{\beta}) f(\underline{b}_i \mid D) d\underline{b}_i \tag{1.18}$$

where $f(y \mid \underline{b}_i, \underline{\beta})$ is the conditional density of $y$ given $\underline{b}_i$ and $f(\underline{b}_i \mid D)$ is the density of random effects which depends on $D$. In general, the maximum likelihood estimates (MLEs) of the parameter $\underline{\beta}$ and $D$, which maximize the likelihood $L(\underline{\beta}, D)$, with respect to $\underline{\beta}$ and $D$, cannot be calculated analytically. Since the integral is intractable especially for high-dimensional random effects, the EM algorithm is one approach to calculate the MLEs iteratively. In the algorithm proposed by Pan and Thompson (1998), the E-step computes the conditional expectation of the log-likelihood

$$Q(\left[\underline{\beta}, D\right] \mid \left[\underline{\beta}', D'\right]) = \int \log f(y, \underline{b}_i \mid \underline{\beta}, D) f(\underline{b}_i \mid y, \underline{\beta}', D') d\underline{b}_i, \tag{1.19}$$

where $f(y, \underline{b}_i \mid \underline{\beta}, D)$ is the joint likelihood of $y$ and $y$ depending on $\underline{\beta}$ and $D$. $f(\underline{b}_i \mid y, \underline{\beta}, D)$ is the posterior density of $\underline{b}_i$ which only depends on a previously fixed value of $\underline{\beta}'$ and $D'$. The estimate of $\underline{\beta}$ and $D$ can be obtained by maximizing $Q(\cdot)$ at the present cycle. They claimed that QMC-EM approach can be viewed as an alternative to a Gauss-Hermite quadrature method, particularly for GLMMs with high dimensional random effects where the Gauss-Hermite quadrature method is less appropriate since the number of integration nodes required increases exponentially with the dimension of the random effects.

In general statistical inferences for GLMMs are greatly hampered by the need for numerical integrations since the integrals involved have no analytical forms in general. Breslow and Clayton (1993) review approaches for solving of the integrals. Many of the methods are quite laborious.

The Gibbs sampling approach to GLMs requires only a minor extension to accommodate the introduction of random effects. A particularly attractive feature is that the amount of computation depends only linearly upon the total number of parameters. This approach is most attractive for a Bayesian formulation of the model.

## 1.2 Bayesian Statistical Modelling

It was mentioned in section 1.1 that the approach taken in this thesis is essentially Bayesian. Therefore we now present the main ideas of Bayesian statistical modelling. Bayesian data analysis had become increasingly popular in 1990's. This is partly due to the widespread use of computers and introduction of Markov chain Monte Carlo methods for statistical inference. A brief description of Markov chain Monte Carlo methods is presented in the next section.

Bayesian statistical data analysis uses probability distributions to make inferences in the form of posterior probability distributions of the unknown model parameters and predictive probability distributions of future events. The major characteristic of Bayesian methods is the use of probability for quantifying uncertainty in inferences. From a Bayesian point of view, there is no distinction between observables and parameters of a statistical model. That is to say all are considered random quantities. The process of Bayesian modelling can be summarized in the following four steps:

1. Building up an appropriate joint probability distribution for observable and unobservable quantities in a problem. The model should be realistic in relation to the underlying scientific problem and to the data collected.

2. Forming the posterior distribution. Let $X$ denote the data set observed and let $\underline{\beta}$ denote the model parameters. Let the joint distribution of $X$ and $\underline{\beta}$ be $P(X, \underline{\beta})$. Then

$$P(X, \underline{\beta}) = P(\underline{\beta}) P(X \mid \underline{\beta}) \tag{1.20}$$

where $P(\underline{\beta})$ is often referred as the prior distribution and $P(X \mid \underline{\beta})$ is the likelihood function, or we can write

$$\text{Full probability model} = \text{Prior distribution} \times \text{Likelihood function}$$

By Bayes theorem,

$$P(\underline{\beta} \mid X) = \frac{P(\underline{\beta})P(X \mid \underline{\beta})}{\int P(\underline{\beta})P(X \mid \underline{\beta})d\underline{\beta}} \tag{1.21}$$

This is called the posterior distribution $\underline{\beta}$, and is the main object of Bayesian inference

3. Evaluation of the final model. It is natural to ask the following questions after a final model is obtained: Does the final model fit the data ? What are the implications of the resulting posterior distribution ? Are the conclusions reasonable ? To answer these questions, one needs to check carefully the final model. If necessary, one can go back to step 1 to alter or expand the model.

4. Inference–: After the probability model is accepted, one can draw inferences about the model parameters and make predictions if necessary about the probabilities of future events using the relevant probability distributions.

(i)Probability Interval: One possible Bayesian inference summary is the $(1-\alpha)100\%$ probability interval, or credible interval, for an unknown quantity of interest. The interval can be regarded as having a certain probability of containing the unknown quantity, in contrast to a frequentist (confidence) interval, which may strictly be interpreted only in relation to a sequence of similar inferences that might be made. Increasing emphasis has been placed on interval estimation rather than hypothesis testing in areas of applied statistics.

(ii) Predictions: Let $\underline{\beta}$ denote the unobservable vector quantities or population parameters of interest, $y$ denote the observed data, and $\tilde{y}$ denote unknown but potentially observable quantities. After the model is accepted, one can calculate the $P(\tilde{y} \mid y)$ which is better known as the posterior predictive distribution where

$$\begin{aligned} P(\tilde{y} \mid y) &= \int P(\tilde{y}, \underline{\beta} \mid y)d\underline{\beta} \\ &= \int P(\tilde{y} \mid \underline{\beta}, y)P(\underline{\beta} \mid y)d\underline{\beta} \\ &= \int P(\tilde{y} \mid \underline{\beta})P(\underline{\beta} \mid y)d\underline{\beta} \end{aligned} \tag{1.22}$$

To illustrate the meaning of $P(\tilde{y} \mid y)$, consider an example of recorded weights of an object weighed n times on a scale. The unknown true weight of the object may be $\mu$ and the true variance is $\sigma^2$. Then $\tilde{y}$ is interpreted as the recorded weight of the object in a planned new weighing.

## 1.3 Markov Chain Monte Carlo Techniques

Markov chain Monte Carlo (MCMC) methodology provides enormous scope for realistic statistical modelling. This method has gained popularity in aiding Bayesian analysis of complex statistical models, since the mid-eighties. MCMC is essentially a Monte Carlo integration method using Markov chains. Bayesians, and sometimes also frequentists, need to integrate over possibly high-dimensional probability distributions to make inferences about model parameters or to make predictions. Bayesians need to integrate over the posterior distribution of model parameters given the data, and frequentists may need to integrate over the distribution of observables given parameter values. Monte Carlo integration draws samples from the required distribution, and then forms sample averages to approximate expectations. The Markov Chain Monte Carlo approach draws these samples by running a cleverly constructed Markov chain for a long time. There are many ways of constructing these chains, but all of them, including the Gibbs sampler (Geman and Geman, 1984), may be thought of as special cases of the general framework of Metropolis *et al.* (1953) and Hastings (1970). Many MCMC algorithms are hybrids or generalizations of the simplest methods: the Gibbs sampler and the Metropolis-Hastings algorithm.

### 1.3.1 The Gibbs sampler

Many statistical applications of MCMC use the Gibbs sampler, which is easy to implement. Gelfand and Smith (1990) gave an overview, and suggested the approach for Bayesian computation. First, as in Smith (1991), we denote probability densities by square brackets, i.e. $[X] = F'(x)$, where $F(x)$ is the cumulative distribution function (CDF) of $X$. Therefore, in the sequel, joint, conditional and marginal densities appear as $[X, Y]$, $[X|Y]$, and $[Y]$ respectively. Now the Gibbs sampling algorithm is best described as follows: Let $X = (X_1, X_2, ..., X_k)$ be a collection of random variables. Given arbitrary initial values ${X_1}^{(0)}, ..., {X_k}^{(0)}$, we draw ${X_1}^{(1)}$ from the conditional distribution $[X_1 \mid {X_2}^{(0)}, ..., {X_k}^{(0)}]$, then ${X_2}^{(1)}$ from $[X_2 \mid {X_1}^{(1)}, {X_3}^{(0)}, ..., {X_k}^{(0)}]$ and so on until ${X_k}^{(1)}$ which comes from $[X_k \mid {X_1}^{(1)}, ..., {X_{k-1}}^{(1)}]$. This scheme is a Markov chain, with equilibrium distribution $[X]$. After $t$ such iterations we would arrive at $({X_1}^{(t)}, {X_2}^{(t)}, ..., {X_k}^{(t)})$. Thus, for $t$ large enough, $X^{(t)}$ can be viewed as a simulated observation from $[X]$. Provided we allow a suitable burn-in time, $X^{(t)}, X^{(t+1)}, X^{(t+2)}, \ldots$ can be thought of as a dependent sample from $[X]$.

Similarly, suppose we wish to estimate the marginal distribution of a variable $Y$ which is a function $g(X_1, X_2, ..., X_k)$ of $X$. Evaluating $g$ at each of the $X^{(t)}$ provides a sample of $Y$. Marginal moments or tail areas are estimated by the corresponding sample quantities. The densities may be estimated using a kernel method.

### 1.3.2 Monitoring Convergence in a Gibbs Sampler Scheme

Markov chain simulation is a very useful tool in model parameter estimation. However there are certain risks of errors if the simulation scheme is not properly monitored in several aspects. The first aspect is the choice of an appropriate model to fit the data. The second is errors in calculation or programming and the last one is slow convergence or convergence to a false target distribution. In monitoring convergence we try to estimate the difference between results based on Markov chain simulation and the desired target distribution. There are two main methods of monitoring convergence in a Markov chain Monte Carlo scheme.

(1) Method of calculating the potential scaling reduction factor (PSRF). This method was proposed by Gelman and Rubin (1992). Their approach is inspired by the method of analysis of variance. It involves forming an overestimate and an underestimate of

the variance of the target distribution with the property that the estimates will be roughly equal at convergence but not before. Suppose we have $m$ parallel Gibbs sampler sequences of length $n$ denoted as $\psi_{ij},\ j = 1, 2, \ldots, n;\ i = 1, 2, \ldots, m$. Let

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{\psi}_{i.} - \bar{\psi}_{..})^2, \quad \text{where } \bar{\psi}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\psi_{ij},\ \bar{\psi}_{..} = \frac{1}{m}\sum_{i=1}^{m}\bar{\psi}_{i.}$$

$$W = \frac{1}{m}\sum_{i=1}^{m}s_i^2, \text{ where } s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\psi_{ij} - \bar{\psi}_{i.})^2$$

The quantity $B$ is the between-sequence variance and $W$ is the within-sequence variance. Then we compute

$$\widehat{var}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B \tag{1.23}$$

where $\widehat{var}(\psi)$ is an unbiased estimate of the variance under stationarity. As $n \longrightarrow \infty$, both $\widehat{var}(\psi)$ and $W$ approach $var(\psi)$, but from opposite directions. Finally, we calculate the quantity $\sqrt{\hat{R}}$ which called the estimated potential scale reduction factor where

$$\sqrt{\hat{R}} = \sqrt{\frac{\widehat{var}(\psi)}{W}} \tag{1.24}$$

If the simulation has converged, $\hat{R}$ converges to a limit of 1.

(2) The second method was proposed by Geyer (1992). This method is to calculate the batch means. It is relatively easy to implement. For each sufficiently long ($n = 10,000$ say) simulated Markov chain, we divide the whole chain into $m$ batches. Each batch has $n$ simulated values and in each batch the sample mean $\bar{x}_i, i = 1, 2, \ldots, m$ is calculated. Then the variance $\hat{\sigma}_B^2$ of the batch means is calculate where

$$\hat{\sigma}_B^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{x}_i - \hat{\mu})^2$$

A measure of the overall mean variance is given by $\hat{\sigma}_B^2$, which accounts for dependence in the chain. This is an easy and attractive method for assessing variability, but care must be taken to choose the size of each batch $n$ large enough for the approximations involved to be valid.

### 1.3.3 Sampling Methods in Gibbs Sampler

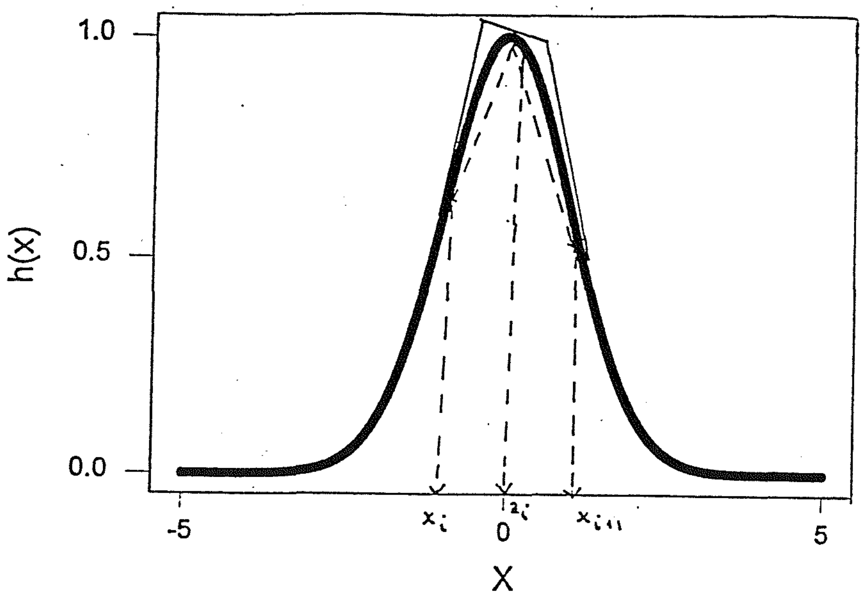
The Gibbs sampler involves sampling from full conditional distributions. It is essential that sampling from full conditional distributions is highly efficient computationally. Rejection sampling and the ratio-of-uniforms are two techniques for sampling independently from a general density $f(x)$ where $f(x)$ is intractable analytically. A third method is the Metropolis-Hastings method as an MCMC method which produces dependent samples.

(1) Rejection sampling method:–Rejection sampling requires an envelope function $G$ of $f$ ($G(x) \geq f(x)$ for all $x$). Samples are drawn from density proportional to $G$, and each sampled point $x$ is subjected to an acceptance/rejection test.

(2) Ratio-of-uniforms method:–Ratio-of-uniforms method is to introduce two variables $U$ and $V$. Let $D$ denote a region in $\{U, V\}$ space defined by $0 \leq U \leq \sqrt{f(V/U)}$. Sample a point $U, V$ uniformly from $D$. This can be done by enveloping the entire region of $D$ by a region $\Delta$. $U$ and $V$ can then be generated by rejection sampling.

(3) Adaptive rejection sampling method:–The main method employed in this thesis is the adaptive rejection sampling (ARS) method developed by Gilks and Wild (1992). In the rejection sampling and ratio-of-uniforms sampling methods, finding a tight envelope function $G$ or a envelope region $\Delta$ is difficult. These can also be very time consuming in the sampling stage. However in many applications of Gibbs sampling, the full conditional densities $f(x)$ are often log-concave (that is $\frac{d^2 ln f(x)}{dx^2} < 0$). In particular, this is true for all generalized linear models with canonical link function (see section 1.1). Gilks and Wild (1992) proposed the adaptive rejection sampling method to sample from a complicated full conditional density which satisfies the log-concavity condition. They showed that an envelope function for $\log f(x)$ can be constructed by drawing tangents to $\log f$ at each abscissae for a given set of abscissae. An envelope between any two adjacent abscissae is then constructed from the tangents at either end of that interval (see Figure 1.1). Secants are drawn through $\log f$ at adjacent abscissae. The envelope is piece-wise exponential for which sampling is straightforward.

Figure 1.1: Graphical representation of the adaptive rejection method




Since generalized linear models for ordered categorical data are considered in this thesis, we will use the adaptive rejection sampling method over the rejection sampling or ratio-of-uniforms method as the ARS is more direct and efficient in terms of sampling from the full conditional density. One may refer to the article by Gilks and Wild (1992) for more theoretical details about their method. Also one may refer to the book by Devroye (1987) or to the book by Gilks *et al.* (1995) for more details on rejection sampling method, ratio-of-uniforms method and the Metropolis-Hastings method.

## 1.4 A Brief Description of Work

In Chapter 2, the generalized linear model with random effects is considered for modelling longitudinal binary data. Zeger *et al.* (1988) presented the modelling longitudinal data using a generalized estimating equation approach (GEE). A Bayesian approach is adopted to deal with this model. Markov chain Monte Carlo techniques are used for parameter estimation. Zeger and Karim (1991) used the Gibbs sampler approach to estimate the model parameters. Their sampling method in the Gibbs sampler scheme is an approximate rejection sampling method. Here we use the adaptive rejection sampling (ARS) method introduced by Gilks and Wild (1992). Simulation results are presented and the method is applied to experimental data from British Telecommunication (BT). Overall simulation results are promising. The concept of latent residuals introduced by Albert and Chib (1995) is used for diagnostic model checking.

Chapter 3 describes the modelling of categorical ordinal response data. The cumulative logit model with random effects is considered here. Again a Bayesian approach is adopted using the concept of latent variables. Gibbs sampling is the main tool for parameter estimation. The cumulative logit model using a logistic link function satisfies the log-concavity conditions. Therefore the adaptive rejection sampling (ARS) method can be used directly. This method is useful because on the one hand it can be used to estimate the regression coefficients in the model, and on the other hand it can also be used to estimate directly the random effects components as well as the latent variables. Results from simulations are promising. The model is also used to analyze the BT experimental data. It is found that random effects are quite large. Latent residuals are also examined at the final stage of model analysis.

In chapter 4 a random effects cumulative logit model with a subject-specific scaling term is developed. The reason is that the response subjects in the British Telecom experiments seem to have their own ordering scale. That is to say, each subject in the experiments tends to have his/her own cut-points in mind. The distribution of response categories show that clusterings occur in the lower and upper ends of the scales. The model discussed in this chapter is largely based on the model developed by Kijewski *et al.* (1989). It is found that the techniques used in chapter 3 can be used again for the new model; but we need to employ a Metropolis-Hastings sampling scheme to generate the random scaling term. This new model is again applied to analyze British Telecom experimental data.

Chapter 5 is devoted to the continuation-ratio logit model, an alternative model for ordinal data. This is a model for the probability of one particular category j given that either category j or higher categories have occured. If the link function is other than the logistic link, we may have the continuation-ratio model with complementary log-log link and continuation-ratio model with probit link when the link function is the complementary log-log and probit respectively. The continuation-ratio logit model is easy to use since any continuation-ratio logit model can be reduced to a binary logistic regression problem for each category j (see Agresti, 1996). If the continuation-ratio link model is considered to have three levels, then each level models one of the binomial probabilities in the expanded likelihood. Therefore the results of chapter 2 for binary regression can be used in this chapter. Analysis of simulated data as well as BT data is presented in this chapter.

In chapter 6 a random effects cumulative probability model with serial correlation is considered. In a time dependent longitudinal study, it will often be necessary to assume that there exists serial correlation amongst the ordinal responses. That is, at any particular time point, the response depends on the previous ones. To model data of this nature, it is easiest to use the probit link instead of logit link, since estimation work is easier to handle if the innovation term of the time series component in the model is driven by a Gaussian distribution. It is found that the ARS method can still be used for parameter estimation as the serial correlation model derived for ordinal response satisfies the log-concave conditions. A simulation study is carried out to verify the method. It is found that a vague prior distribution should be used rather than a non-informative prior distribution in order to obtain stable results for the model parameters. Finally chapter 7 contains some concluding remarks.

# Chapter 2

# A Gibbs Sampling Approach for Modelling repeated Binary Data

## 2.1 Introduction

In this chapter a random effects generalized linear model is considered for modelling binary data. This serves as a starting point for more complicated models for ordered categorical data in later chapters. Perhaps the most widely used of generalized linear models are those for binary or binomial data. Suppose that observation $y$ follows the Bernoulli distribution with parameter $p$. If one chooses the logit transformation of the probability of success, $h(p) = \log\left(\frac{p}{1-p}\right)$ as the link function, then the resulting generalized linear model will be the logistic regression model. Other link functions are often used. In particular the probit link, $h(\mu) = \Phi^{-1}(\mu)$, is another popular choice. The probit model is commonly used in economic and social sciences. In practice, the probit and logit models are quite similar, differing mainly in the extreme of the tails. One advantage of the logit model is one can work out the log-odds ratio comparing two covariate patterns given the regression coefficient vector $\underline{\beta}$. The log-odds ratio is interpreted as the relative odds of success for one covariate pattern over the other. This advantage does not apply to the probit model. Further details can be found in the article by McCullagh (1980).

In the following consider the random effects generalized linear model with logit link function for repeated measures on binary data. By repeated measures, we mean that binary observations $\{y_{ij}\}$, subjects, $i = 1, 2, \ldots, n$, occasions, $j = 1, 2, \ldots, k$ are obtained. The logit model with random effects is given by,

$$\text{logit}\left(\frac{p_{ij}}{1-p_{ij}}\right) = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i, \tag{2.1}$$

where

(i) $\underline{x}_{ij}$: $p \times 1$ vector of covariates associated with response $y_{ij}$,

(ii) $\underline{\beta}$: $p \times 1$ vector of regression coefficients,

(iii) $\underline{z}_{ij}$: $q \times 1$ vector of random effects component. $\underline{z}_{ij}$ is a subset of $\underline{x}_{ij}$,

(iv) $\underline{b}_i$: $q \times 1$ vector of coefficients of the random effects component.

In this chapter a Bayesian approach is taken to model equation 2.1. The Gibbs sampler is employed to estimate the regression coefficient vector $\underline{\beta}$ and the variance of the random effects component $Var(\underline{b}_i)$.

## 2.2 Generalized Linear Model with Random Effects in Bayesian Setting

We assume that $\underline{b}_i$ follows a multivariate normal distribution with mean 0 and variance-covariance $D$ and the likelihood function for the parameter vector $\underline{\beta}$ and $D$ has the form

$$L(\underline{\beta}, D, y) \propto \prod_{i=1}^{n} \int \prod_{j=1}^{k} f(y_{ij} \mid \underline{b}_i) | D |^{-\frac{1}{2}} \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i) d\underline{b}_i \tag{2.2}$$

Further let $P(\underline{\beta}, D)$ be the joint prior distribution for $\underline{\beta}$ and D, then the posterior distribution $P_1(\underline{\beta}, D \mid y)$ is given by

$$P_1(\underline{\beta}, D \mid y) = \frac{\prod_{i=1}^{n} \int \prod_{j=1}^{k} f(y_{ij} \mid \underline{b}_i, \underline{\beta}) g(\underline{b}_i \mid D) P(\underline{\beta}, D) d\underline{b}_i}{\int \prod_{i=1}^{n} \int \prod_{j=1}^{k} f(y_{ij} \mid \underline{b}_i, \underline{\beta}) g(\underline{b}_i \mid D) P(\underline{\beta}, D) d\underline{b}_i dD d\underline{\beta}} \tag{2.3}$$

where $g(\underline{b}_i \mid D)$ is the Gaussian density of $\underline{b}_i$. The marginal posterior densities of $\underline{\beta}$ and $D$ can be obtained from the joint posterior distribution in equation 2.3 by integrating out $D$ and $\underline{\beta}$ respectively. We can also derive important Bayesian summaries from the numerator alone since the denominator is a normalizing constant independent of $\underline{\beta}$ and $D$.

### 2.2.1 Conditional Posterior Densities

To derive the conditional posterior densities, we consider the full posterior density, including the random effects i.e.

$$P_1(\underline{\beta}, \underline{b}_i, D) \quad \propto \prod_{i=1}^{n}\prod_{j=1}^{n} f(y_{ij} \mid \underline{b}_i, \underline{\beta}) g(\underline{b}_i \mid D) P(\underline{\beta}, D) \tag{2.4}$$

where $\underline{b}_i \mid D \sim N(0, D)$ and $f(y_{ij} \mid \underline{b}_i) = \exp\{[y_{ij}\theta_{ij} - a(\theta_{ij}) + c(y_{ij})]/\phi\}$. Now for the logistic model, let $y_{ij}$ follow Bernoulli distribution with parameter $p_{ij}$ where $0 < p_{ij} < 1$, i.e.

$$f(y_{ij}) = {p_{ij}}^{y_{ij}}(1 - p_{ij})^{1-y_{ij}}, \qquad y_{ij} = 0 \quad or \quad 1.$$

Also

$$\begin{aligned} f(y_{ij}) &= \frac{e^{\theta_{ij} y_{ij}}}{1 + e^{\theta_{ij}}} \\ \theta_{ij} &= \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i \end{aligned}$$

and $p_{ij} = \frac{e^{\theta_{ij}}}{1+e^{\theta_{ij}}}$. Therefore the posterior density is given by (up to proportionality)

$$\begin{aligned} P_1(\underline{\beta}, \underline{b}_i, D) = \quad & c \times \prod_{i=1}^{n}\prod_{j=1}^{k} \frac{e^{(\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i)y_{ij}}}{1 + e^{\underline{x}'_{ij}\underline{\beta} + \underline{x}'_{ij}\underline{b}_i}} \mid D \mid^{-(1/2)} \\ & \times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1} \underline{b}_i) P(\underline{\beta}) P(D) \end{aligned} \tag{2.5}$$

It is assumed that $P(\underline{\beta}, D) = P(\underline{\beta})P(D)$ in equation 2.3. If standard uniform non-informative priors are used for $P(\underline{\beta}, D)$, i.e. $P(\underline{\beta}, D) \propto$ constant, then

$$\begin{aligned} P_1(\underline{\beta}, \underline{b}_i, D) \quad & = c_2 \times \prod_{i=1}^{n}\prod_{j=1}^{k} \frac{e^{(\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i)y_{ij}}}{1 + e^{\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i}} \mid D \mid^{-(1/2)} \\ & \times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1} \underline{b}_i) \end{aligned} \tag{2.6}$$

where $c_2$ is also a constant and $P_1(\underline{\beta}, \underline{b}_i, D)$ is intractable analytically.

It can be shown that the conditional posterior density of $\underline{\beta}$ is

$$[\underline{\beta} \mid D, \underline{b}] \propto \prod_{i=1}^{n}\prod_{j=1}^{k} \frac{e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)y_{ij}}}{1+e^{\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i}} \tag{2.7}$$

i.e. $[\underline{\beta} \mid D, \underline{b}_i] = [\underline{\beta} \mid \underline{b}_i]$, which is independent of $D$.

Similarly the conditional posterior density of $\underline{b}_i$ is

$$[\underline{b}_i \mid D, \underline{\beta}] \propto \prod_{j=1}^{k} \frac{e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)y_{ij}}}{1+e^{\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i}} \times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i) \tag{2.8}$$

i.e. $[\underline{b}_i \mid \cdot] = [\underline{b}_i \mid \underline{\beta}, D]$, which is a function of $\underline{\beta}$ and $D$. Further the conditional posterior density of $D$ is

$$[D \mid \underline{\beta}, \underline{b}_i] \propto \times \mid D \mid^{-(1/2)} \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i) \tag{2.9}$$

i.e. $[D \mid \underline{\beta}, \underline{b}_i] \propto [D \mid \underline{b}_i]$, which depends only on $\underline{b}_i$.

### 2.2.2 Log-concavity Conditions

In order to use adaptive rejection sampling method to carry-out Gibbs sampling, we need first to check carefully whether the conditional posterior densities satisfy the log-concavity conditions. That is, for a conditional density for $\beta_v$, $h$ (say), we check whether $\frac{\partial^2 \ln h}{\partial^2 \beta_v} < 0$ and for a conditional density for $b_v$ we check whether $\frac{\partial^2 \ln h}{\partial^2 b_v} < 0$. In fact we can prove that, in general, the conditional posterior distribution $h(\underline{\beta})$ is log-concave with respect to $\underline{\beta}$ and $h(\underline{b}_i)$ is log-concave with respect to $\underline{b}_i$ for any design matrix. The proof is stated as follows.

**Proof:**

Consider the log-conditional posterior distribution of $\underline{\beta}$, that is,

$$\ln h(\underline{\beta}) = c' + \sum_{i=1}^{n}\sum_{j=1}^{k}\left[(\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i)y_{ij} - \ln(1+e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)})\right]; \qquad c' \text{ is a constant.}$$

To show this function is concave with respect to the coefficient vector $\underline{\beta}$, it suffices to show that $\ln(1+e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)})$ is convex with respect to $\underline{\beta}$ for each combination of $i$

and $j$; since $\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$ is a linear function with respect to $\underline{\beta}$ which is concave. Now consider $\ln(1 + e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)})$, let

$$F_1(\underline{\beta}) = \ln(1 + e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}).$$

Then

$$\begin{aligned} \frac{\partial F_1}{\partial \underline{\beta}} &= \left[\frac{e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}}{1 + e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}}\right]\underline{x}_{ij} \\ &= u_{ij}\underline{x}_{ij} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 F_1}{\partial \underline{\beta}\partial \underline{\beta}'} &= u_{ij}(1 - u_{ij})\underline{x}_{ij}\underline{x}'_{ij} \\ &= (1 - u_{ij})u_{ij}\mathbf{X} \end{aligned}$$

where $u_{ij} = \left[\frac{e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}}{(1+e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)})}\right]$ and $\mathbf{X} = \underline{x}_{ij}\underline{x}'_{ij}$ is a p×p symmetric matrix. It is straightforward to see that $0 < u_{ij} < 1$ and the symmetric matrix $\mathbf{X}$ is non-negative definite. Therefore $F_1$ is convex.

Similarly, to show that the marginal posterior distribution of $\underline{b}_i$ is log-concave with respect to $\underline{b}_i$, let's consider equation 2.13 and taking the natural logarithm, we have

$$\ln h(\underline{b}_i) = k' + \sum_{j=1}^{k}(\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i)y_{ij} - \sum_{j=1}^{k}\ln(1 + e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}) - \frac{1}{2}\underline{b}'_i D^{-1}\underline{b}_i,$$

where $k'$ is a constant. Let $F_2 = \ln(1 + e^{(\underline{x}'_{ij}\underline{\beta}+\underline{z}'_{ij}\underline{b}_i)}) + \frac{1}{2}\underline{b}'_i D^{-1}\underline{b}_i$. To show that $h(\underline{b}_i)$ is log-concave with respect to $\underline{b}_i$, it suffices to show that $F_2$ is convex for fixed $i$ and $j$. Along the above lines, it can be shown that

$$\begin{aligned} \frac{\partial^2 F_2}{\partial \underline{b}_i \partial \underline{b}'_i} &= (1 - u_{ij})u_{ij}\underline{z}_{ij}\underline{z}'_{ij} + D^{-1} \\ &= (1 - u_{ij})u_{ij}\mathbf{Z} + D^{-1} \end{aligned}$$

where $\mathbf{Z} = \underline{z}_{ij}\underline{z}'_{ij}$ is a q×q symmetric matrix. The symmetric matrix $\mathbf{Z}$ and the inverse of variance-covariance $D$ are non-negative definite. Therefore $F_2$ is also convex. This completes the proof of log-concavity condition for any design matrix of the model. The above results still hold if we include any log-concave prior for $\underline{\beta}$.

In fact, Wedderburn (1976) showed that generalized linear models with many link functions , such as normal, logit, probit, Poisson and complementary log-log links, in the exponential family satisfy the concavity condition.

As the conditional posterior densities for $\underline{\beta}$ and $\underline{b}_i$ satisfy the log-concavity condition. We can employ adaptive rejection sampling method to carry-out Gibbs sampling in the following simulation work. Generation of $D$ can be obtained by using the algorithm of Odell and Feiveson (1966).

## 2.3 Simulation Studies

Zeger and Karim (1991) considered the following logistic regression model

$$\begin{aligned} \text{logit}P(y_{ij} = 1 \mid b_i) &= \beta_0 + \beta_1 t_j + \beta_2 x + \beta_3 t_j x + b_{0i} + b_{1i} t_j \\ &= \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i \end{aligned} \tag{2.10}$$

where $\underline{x}'_{ij} = (1, t_j, x_i, t_j x_i), \underline{z}'_{ij} = (1, t_j), \underline{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $\underline{b}' = (b_{0i}, b_{1i})$. In their simulation study, Zeger and Karim (1991) put $x_i = 0$ for half the population and 1 for the remainder and $t = -3, -2, -1, 0, 1, 2, 3$ for each subject. Each data set comprised of $I = 100$ clusters of size $n_i = 7$. The fixed effects coefficients were set at $\beta_0 = -2.5, \beta_1 = +1.0, \beta_2 = -1.0$, and $\beta_3 = +0.5$. The same set of figures is used here for the adaptive rejection sampling scheme and our random effects distribution was simulated with $E(\underline{b}_i) = 0$ and $Var(\underline{b}_i) = D$ where

$$D = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.001 \end{bmatrix}. \tag{2.11}$$

The regression and random effect parameters can be generated solely using the ARS method since each of the conditional posterior distributions of the model parameters satisfy the log-concavity condition. In Zeger and Karim's (1991) paper, their methods for carrying out the Gibbs sampler were rather cumbersome and time consuming. They use ordinary rejection sampling method to simulate the regression coefficient vector $\underline{\beta}$. This part creates no problem, but generating the random effects $\underline{b}_i$ from the conditional posterior distribution $h(\underline{b}_i \mid \underline{\beta}, D, y)$ is the most time consuming step. Since

$$h(\underline{b}_i \mid \underline{\beta}, D, y) \propto h(y \mid \underline{b}_i, \underline{\beta}) g(\underline{b}_i \mid D) P(\underline{\beta}, D)$$

Denote $h(\underline{b}_i \mid \cdot)$ as $p(\underline{b}_i)$. To sample a value from $p(\underline{b}_i)$ using rejection sampling, it is necessary to find the mode and curvature of $p(\underline{b}_i)$. Finding the mode of $p(\underline{b}_i)$ involves solving a ridged regression by iterative weighted least squares and curvature must be obtained by iterating the equations for maximum value $\hat{\underline{b}}_i$ of $p(\underline{b}_i)$ and its curvature $\hat{\underline{v}}_i$ where

$$\hat{\underline{b}}_i = (Z_i' V_i Z_i + D^{-1}) Z_i' V_i (y_i - X_i \underline{\beta}) \tag{2.12}$$

$$\hat{\underline{v}}_i = (Z_i' V_i Z_i + D^{-1})^{-1}, \quad V_i = Var(\hat{\underline{\beta}}_i), \quad Z_i = Var(\underline{b}_i) \tag{2.13}$$

One can refer to Section 5 of the paper by Zeger and Karim (1991) for more details. For the ARS approach, model parameters can be generated iteratively at one-go without employing any optimization techniques; e.g. to locate the mode and curvature of the parameter distributions of $b_{0i}$ and $b_{1i}$. This is the major advantage of ARS method over the ordinary rejection sampling.

### 2.3.1 Simulation Results

To perform Gibbs sampling, we run the chain for each parameter of interest for 11,000 times and discard the initial 1000 values as burn-in. Table (2.1) shows the overall results of one typical run using one simulated data set.

Table 2.1: Simulation results of ARS method

| Parameters | True Value | Mean | Median | Std. Dev. |
|---|---|---|---|---|
| $\beta_0$ | -2.5 | -2.6236 | -2.6031 | 0.3573 |
| $\beta_1$ | +1.0 | 1.1178 | 1.1089 | 0.1624 |
| $\beta_2$ | -1.0 | -1.1251 | -1.0351 | 0.6028 |
| $\beta_3$ | +0.5 | 0.4632 | 0.4232 | 0.3815 |
| $Var(b_0)$ | +1.0 | 1.0879 | 0.9804 | 0.3293 |

In Table 2.1, estimates using the ARS method are generally close to the true values. To monitor convergence, we use the method of batching for each individual series and we also calculated the estimated potential scale reduction factor (PSRF) $\hat{\text{R}}$. In this simulation case $\hat{\text{R}} = 1.0113$ for the regression parameters which indicates that the Gibbs sampler scheme converged very well. These results can be compared with those obtained by Zeger and Karim (1991) which are shown in Table 2.2 below.

Table 2.2: Simulation results of Zeger and Karim

| Parameters | True Value | mean | Std. Dev. |
|---|---|---|---|
| $\beta_0$ | -2.5 | -2.67 | 0.36 |
| $\beta_1$ | +1.0 | +1.07 | 0.15 |
| $\beta_2$ | -1.0 | -0.96 | 0.56 |
| $\beta_3$ | +0.5 | 0.49 | 0.24 |
| $Var(b_0)$ | +1.0 | 1.21 | 0.60 |

As we can see in Table 2.1, estimates using the ARS method are generally close to the true values. Our results are similar to those obtained by Zeger and Karim (1991). However, our method is much more efficient.

## 2.4 Telephone Connection Quality Data

At the laboratories of British Telecom (BT) in Martlesham near Ipswich (UK), a series of experiments concerning the quality of telephone connections were conducted. One of their experiments is called conversation experiment. A conversation experiment consists of a number of pairs of subjects and each pair engages in a conversation over the telephone. The two subjects in each pair sit in two different cabinets; say cabinet A and B. In a conversation experiment a subject engages in conversation and then gives an opinion about the telephone connection. The duration of conversation is determined by the subjects. When the conversation is finished the subjects hang up and are prompted by the experiment controller to give an opinion of the transmission condition. The opinion is typically given on a five point scale graded from 'Bad' to 'Excellent'. This is an ordinal response scale. The subjects also give a binary responses to a question on difficulty in hearing over the connection. In this chapter we use the methodology described in Section (1) and (2) to analyze the dependence between the factors and the binary response. The analysis of ordinal responses will be treated in next chapter.

The order in which a subject hears the transmission conditions is determined by an experimental design. This design can be set out as a two-way layout in which each row corresponds to a subject and each column corresponds to a period. In each period there is a particular level of transmission conditions. The logit model is linear in two factors; namely (i) rows (random effects), and (ii) transmission conditions.

For the experiments that we analyzed, one is called E199 experiment. This experiment has an unlimited duration in the conversation between the two subjects. There are 2 pairs of 16 subjects. Each subject (row) received 8 trials. In each trial one level of the 8 transmission conditions is set. Altogether, we have 1 + 32 + (8-1)=40 parameters to be estimated as well as the random effect variance. The model is represented as follows:

$$\text{logit}P(y_{ij} = 1 \mid b_{0i}) = \beta_0 + \beta_{33}\text{I}[\text{cond. } 2] + \cdots + \beta_{39}\text{I}[\text{cond. } 8] + b_{0i} \qquad (2.14)$$

where $\text{I}[\cdot]$ is the indicator variable. In the following we present the final results of our analysis. As in the simulation, we generated 10,000 random variates after 1000 burn-in values for each parameter. In Table 2.3 we show the results of our estimates for experiment E199.

Table 2.3. Results of the BT (E199) experiment. The estimates are reported along with the mean, the standard error, the standard error of batching mean, the lower 2.5$th$, ($P_{0.025}$) and the upper 97.5$th$, ($P_{0.975}$) percentiles.

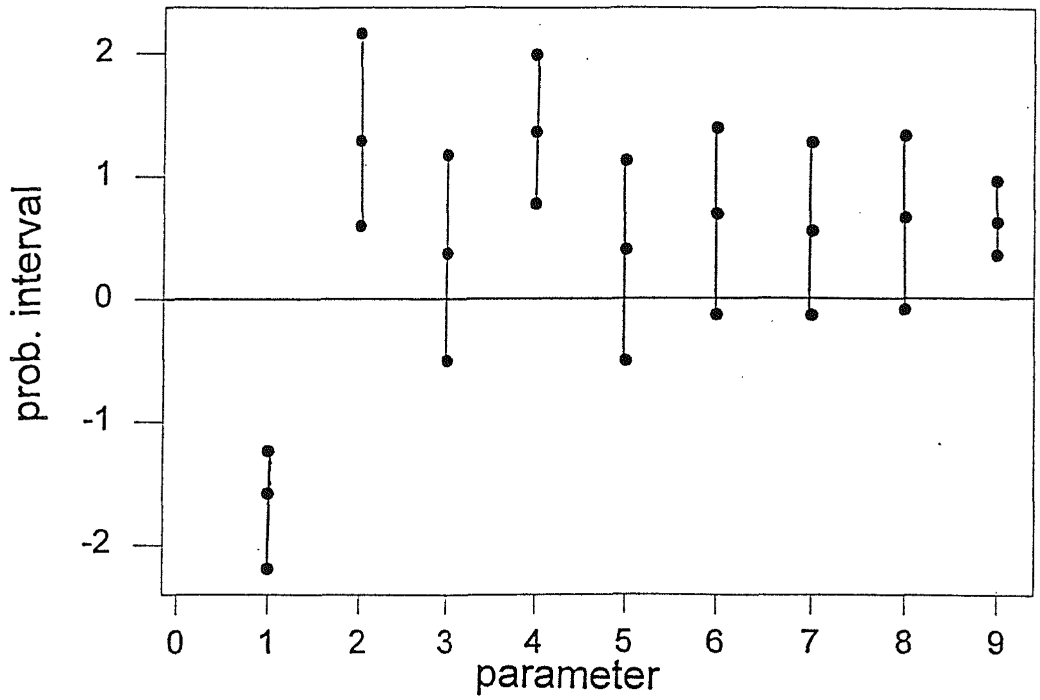
| para. | mean | s.d. | s.d.[1] | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|---|
| intercept | -1.5772 | 0.5026 | 0.0126 | -2.1882 | -1.2311 |
| Cond 1 | 0 | (aliased) | | | |
| Cond 2 | 1.2916 | 0.6412 | 0.0440 | 0.5874 | 2.1638 |
| Cond 3 | 0.3734 | 0.6418 | 0.0505 | -0.5007 | 1.1730 |
| Cond 4 | 1.3627 | 0.5875 | 0.0553 | 0.7773 | 1.9897 |
| Cond 5 | 0.4098 | 0.6289 | 0.0455 | -0.4958 | 1.1329 |
| Cond 6 | 0.7003 | 0.6060 | 0.0530 | -0.1223 | 1.3977 |
| Cond 7 | 0.5600 | 0.6131 | 0.0551 | -0.1232 | 1.2833 |
| Cond 8 | 0.6741 | 0.6376 | 0.0516 | -0.0735 | 1.3408 |
| $\sigma_0^2$ | 0.6369 | 0.3912 | 0.0762 | 0.3724 | 0.9732 |

1: This is the standard deviation of the batching means.

Figures in Table 2.3 show that the effects of condition 2 and 4 are significantly different from level 1. The rest of the conditions have insignificant effects (the 95% probability intervals contain zero values). The mean random effects variance is equal to 0.6369. The batching variances are small relatively to the overall sample variances in each chain. The number of sample points in each batch is 500. There are 20 batches.

The PSRF ($\hat{R}$) value for the regression parameters is equal to 1.0238 for Table 2.3. These figures show that convergence is good in each of the Gibbs sampling scheme. For each parameter, we have a MCMC sample from the posterior distribution of that parameter. A $100(1-\alpha)\%$ probability interval for a parameter may be estimated by taking any range which contains $100(1-\alpha)\%$ of the MCMC sample. As usual, we take the $100\frac{\alpha}{2}\%$ and $100(1-\frac{\alpha}{2})\%$ quantiles of the sample, as the endpoints of the interval. The last two columns in each of the tables show the 2.5% and 97.5% quantile values. The probability intervals and the estimated mean values of the parameter in Table 2.3 are presented in Figure 2.1

Figure 2.1 : Plots of Probability Intervals of Parameters




The empirical posterior distributions of the intercept term, effects of condition 2, condition 8 and random effect variance are shown the following graphs. The sample size in each histogram is equal to 2,000. These values are taken from the final portion of the original simulated chain (n=10,000). A sample of 2,000 is used to summarize the true posterior distribution.

Figure 2.2: Empirical Posterior

Distribution of Intercept

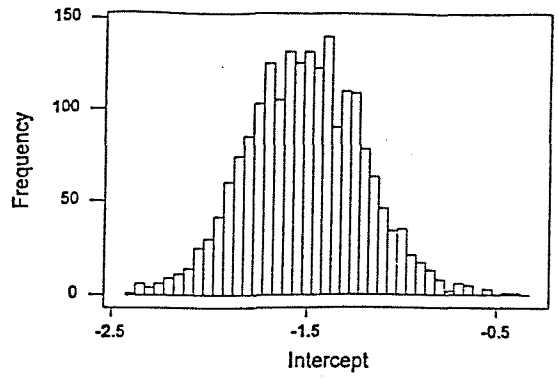



Figure 2.3: Empirical Posterior

Distribution of Condition 2

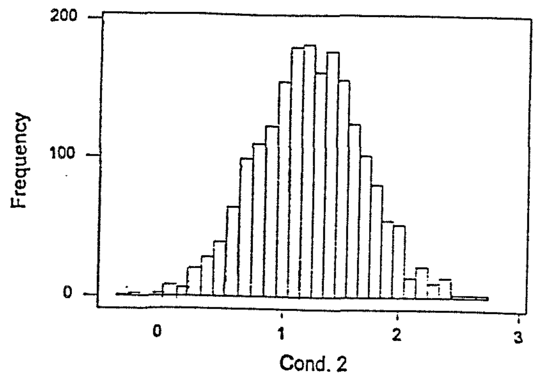



Figure 2.4: Empirical Posterior

Distribution of Condition 8

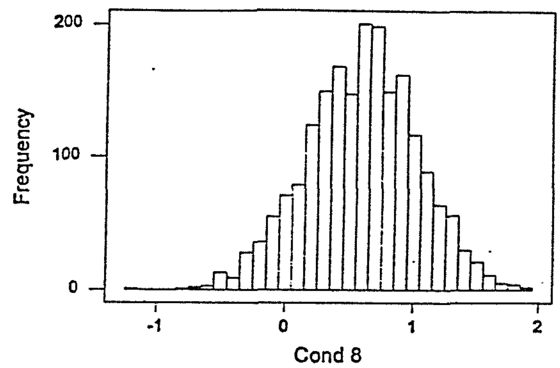
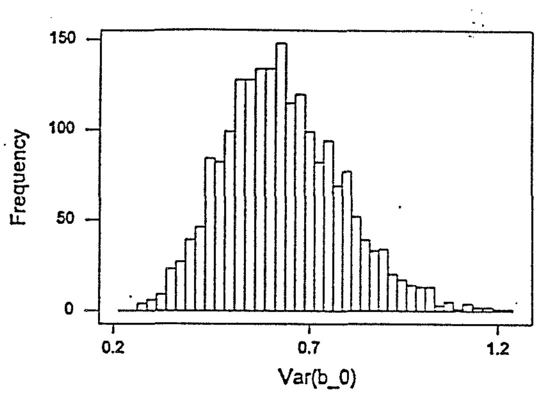



Figure 2.5: Empirical Posterior

Distribution of Random Effect Variance




### 2.4.1 Posterior-Predictive Distributions

One of the advantage of Bayesian statistics is the power of predicting future quantities based on the joint probability model. For this model, the predictive distribution is the probability of $y^*$ successes in a future experiment. The predictive distribution of the future observations for given $\underset{\sim}{\beta}$ is given by

$$f(y_i^* \mid y) = \int f(y_i^* \mid x_i, \underset{\sim}{\beta}) g(\underset{\sim}{\beta} \mid y) d\underset{\sim}{\beta},$$

where for an 'average' individual $f(y_i^*)$ is the binomial likelihood function and $g(\underset{\sim}{\beta} \mid y)$ is the posterior density of $\underset{\sim}{\beta}$.

In practice, the posterior-predictive distribution can be computed via simulation. First we simulate $\underset{\sim}{\beta}^{(j)}, j = 1, 2, \ldots, m$, and at each $j$ we draw a binomial random variable $y_i^*(j)$ with sample size $n_i$ and success probability $p_i^{(j)} = F(\underline{x}_i{}'\underline{\beta}^{(j)})$. A histogram estimate for the posterior-predictive distribution for each binomial observation can be constructed.

For example in the E199 experiment, we compute the posterior-predictive distribution for subject 8 and subject 16 at each condition level. There are eight conditions. At each condition $k$, the observation is binary ($n_k = 1$ for each k). Here the predictive distribution is given by $E(p \mid \text{Condition})$, the posterior mean for $p$. The following Tables show the posterior- predictive distribution of subject 8 and subject 16. These figures are calculated based on the simulated values of $\underline{\beta}^{(j)}, j = 1, 2, \ldots, m$. The sample size $m$ is also equal to 2,000, taken from the last 2,000 values of the entire simulated Markov chain.

Table 2.4. Results of Posterior-predictive distribution of subject 8 for given conditions (E199 experiment).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s. d. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.2251 | 0.3257 | 0.0690 | 0.4678 |
| Cond 2 | 0.4227 | 0.3843 | 0.1643 | 0.7183 |
| Cond 3 | 0.3077 | 0.3601 | 0.1019 | 0.6041 |
| Cond 4 | 0.4506 | 0.3898 | 0.1796 | 0.7485 |
| Cond 5 | 0.2715 | 0.3516 | 0.0855 | 0.5574 |
| Cond 6 | 0.3721 | 0.3748 | 0.1372 | 0.6706 |
| Cond 7 | 0.3244 | 0.3638 | 0.1185 | 0.6166 |
| Cond 8 | 0.3350 | 0.3679 | 0.1187 | 0.6333 |

Table 2.5. Results of Posterior-predictive distribution of subject 16 for given conditions (E199 experiment).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.3384 | 0.3694 | 0.1093 | 0.6439 |
| Cond 2 | 0.5587 | 0.3976 | 0.2394 | 0.8479 |
| Cond 3 | 0.4375 | 0.3895 | 0.1639 | 0.7368 |
| Cond 4 | 0.5868 | 0.3907 | 0.2778 | 0.8565 |
| Cond 5 | 0.3939 | 0.3852 | 0.1426 | 0.7034 |
| Cond 6 | 0.5075 | 0.3921 | 0.2141 | 0.7896 |
| Cond 7 | 0.4557 | 0.3896 | 0.1830 | 0.7520 |
| Cond 8 | 0.4679 | 0.3923 | 0.1812 | 0.7718 |

Figures in Table 2.4 reveal that given all the conditions, subject 8 would have on the average 30% of the chance to give a positive response of having difficulty in hearing from the telephone connection. In Table 2.5, subject 16 would have a over 50% of the chance of giving a positive response that he is having difficulty in hearing given transmission condition 2 and 4. Transmission condition 2 and 4 are identified as having significant positive effects. This is the advantage of conducting a Bayesian statistical analysis. Also this is an important piece of information as far as the telecommunication engineers are concerned.

Also we can get a predictive value for a 'random future subject' by simulating the random effect for that subject. Table 2.6 shows the predictive subject effects for an alternative sample of 32 random future subjects.

Table 2.6. Results of Posterior-predictive distribution of future subject effects (E199 experiment).

| Subject | Pred. subject effect | s.d. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Subject 1 | -0.2919 | 0.7570 | -1.3736 | 0.8469 |
| Subject 2 | -0.3200 | 0.7576 | -1.5072 | 0.7668 |
| Subject 3 | -0.2135 | 0.7758 | -1.3665 | 0.9646 |
| Subject 4 | -0.5640 | 0.7661 | -1.7075 | 0.5518 |
| Subject 5 | 0.7976 | 0.8024 | -0.4220 | 2.0981 |
| Subject 6 | -0.3876 | 0.7486 | -1.5035 | 0.7449 |
| Subject 7 | 0.0384 | 0.7673 | -1.1204 | 1.1830 |
| Subject 8 | -0.0282 | 0.7724 | -1.2289 | 1.1210 |
| Subject 9 | -0.0153 | 0.7366 | -1.1007 | 1.0673 |
| Subject 10 | -0.2683 | 0.7478 | -1.3240 | 0.8145 |
| Subject 11 | 0.7547 | 0.7861 | 0.1261 | 2.0035 |
| Subject 12 | 0.4624 | 0.7432 | -0.5981 | 1.5673 |
| Subject 13 | -0.7922 | 0.7910 | -1.9874 | -0.4045 |
| Subject 14 | -0.0361 | 0.8598 | -1.4648 | 1.4347 |
| Subject 15 | 0.6441 | 0.7888 | -0.5771 | 1.8393 |
| Subject 16 | 0.4769 | 0.7745 | -0.6939 | 1.6608 |
| Subject 17 | -1.0879 | 0.8651 | -2.6257 | -0.3366 |
| Subject 18 | 0.7926 | 0.7573 | 0.1314 | 1.9138 |
| Subject 19 | 0.7572 | 0.7652 | -0.3515 | 1.9092 |
| Subject 20 | -0.4979 | 0.7588 | -1.6085 | 0.5976 |
| Subject 21 | 0.6597 | 0.8233 | -0.7031 | 2.0403 |
| Subject 22 | -0.4202 | 0.7367 | -1.4447 | 0.6355 |

Table 2.6. Results of Posterior-predictive distribution of future subject effects (E199 experiment) (Continued).

| Subject | Pred. subject effect | s.d. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Subject 23 | -0.1931 | 0.7688 | -1.4083 | 0.9539 |
| Subject 24 | 0.1677 | 0.7965 | -1.0904 | 1.3842 |
| Subject 25 | -0.4903 | 0.7642 | -1.5959 | 0.6956 |
| Subject 26 | -0.7817 | 0.7683 | -1.9186 | -0.1170 |
| Subject 27 | 0.4511 | 0.7714 | -0.7355 | 1.6562 |
| Subject 28 | -0.4331 | 0.8879 | -2.0609 | 0.9644 |
| Subject 29 | 0.3446 | 0.7543 | -0.7866 | 1.4492 |
| Subject 30 | 0.2762 | 0.7877 | -0.9538 | 1.5049 |
| Subject 31 | 0.4033 | 0.7694 | -0.7160 | 1.5551 |
| Subject 32 | 0.3846 | 0.7689 | -0.7960 | 1.5403 |

### 2.4.2 Model Checking Using Latent Residuals

Finally for model checking, we use the method of latent residuals introduced by Albert and Chib (1995). That is, if $\operatorname{logit}\Pr(y_{ij} = 1 \mid b_0) = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$ is the correct model for the data, then we can express this model as

$$y_{ij} = \mathrm{I}\left[\lambda_{ij} > 0\right], \tag{2.15}$$

where $\lambda_{ij}$ are the latent variables and I is the indicator variable. $\lambda_{ij}$ can be generated directly from a logistic distribution for given values of $\underline{\beta}^{(t)}$ and $\underline{b_i}^{(t)}$ at each stage of iteration of our Gibbs sampler scheme. $\lambda_{ij}$ are positive if observation $y_{ij}$ is 1 and negative if observation $y_{ij}$ is 0. Therefore latent residuals $\varepsilon_{ij}$ are defined as

$$\varepsilon_{ij} = \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b_i} \tag{2.16}$$

Latent residuals $\varepsilon_{ij}$ are a priori a random sample from a standard logistic distribution. In our studies, a sample size of 10,000 mean latent residuals have been generated. To examine whether or not the generated latent residuals follow a standard logistic distribution, a Kolmogorov-Smirnov test is used. To conduct the test a random sample of size 30 is being picked from the 10,000 generated latent residuals each time. We obtain the necessary $D_n$ test statistic as well as $\Pr(D \leq D_n)$. The test is repeated for 50, 100 and 1000 times. Table 2.7 shows the overall results for cabinet A.

Table 2.7: Results of K-S Test of latent residuals of E199 experiment

| Sample size (n) | No. of runs (m) | Mean $D_n$ | Mean $\Pr(D \leq D_n)$ |
|---|---|---|---|
| 30 | 50 | 0.1517 | 0.6021 |
| 30 | 100 | 0.1502 | 0.5029 |
| 30 | 1000 | 0.1487 | 0.5109 |

The following histogram also show the distribution of the mean latent residuals and the probability plot of the mean latent residuals. The mean latent residuals are obtained at each simulation run. The last 2,000 values of the entire 10,000 are used for plotting the histograms.

Figure 2.6: Histogram of the mean residuals superimposed by a standard logistic curve

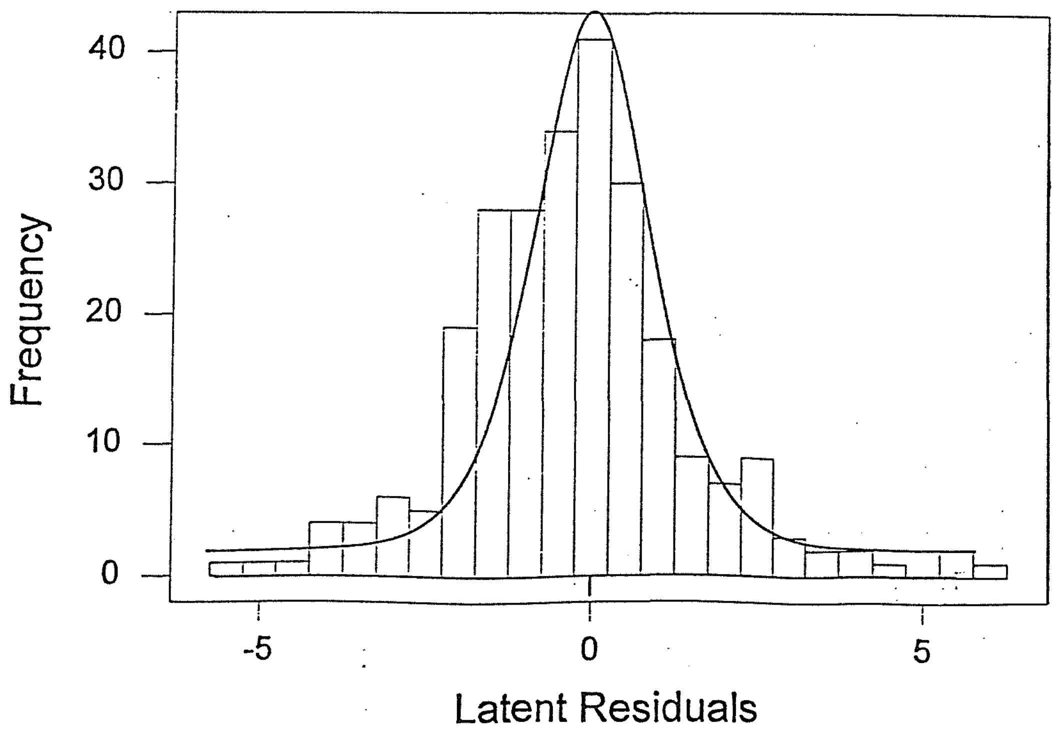
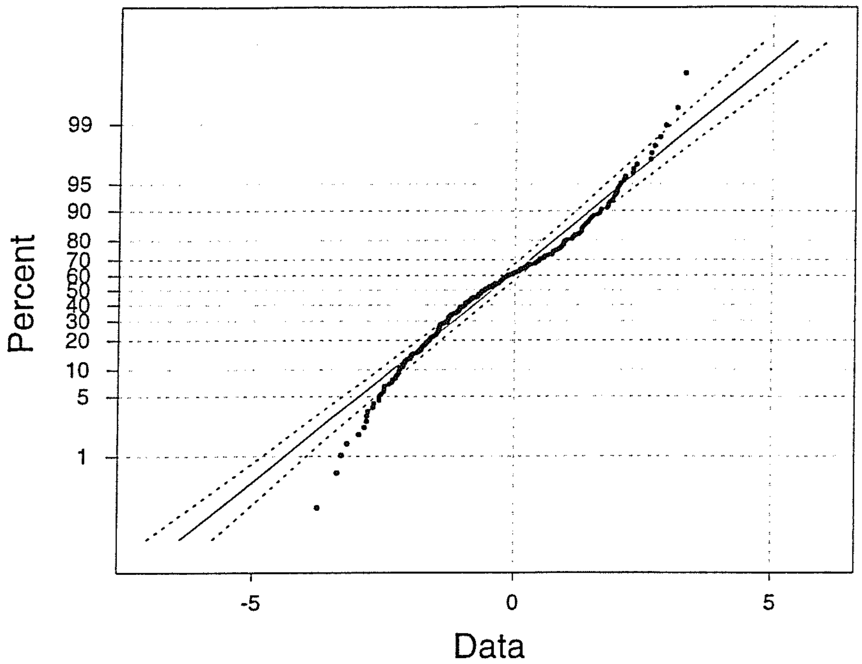


Figure 2.7: Logistic Probability Plot of Mean Latent Residuals




As we can see from the above logistic probability plot, the majority of the mean latent residuals are close to the straight line. This agrees with the histogram of the in Figure 2.6 which is superimposed by a standard logistic curve. The outliers mainly come from subject 7 and subject 23 where model fits for these two subjects are consistently bad. One difficulty in interpreting the posterior mean latent residuals is the loss of identification of the observations in the averaging process. The smallest and largest latent residual may vary over iterations. The smallest and largest residual did not always correspond to the same observations. The use of latent residuals is one of the tools in Bayesian model diagnostic checking. Another way to check the final model is to compute the so-called Bayesian residuals (Albert and Chib, 1993). For logit model, Bayesian residuals are defined as

$$\begin{aligned} r_i &= \frac{y_i}{n_i} - p_i \\ &= \hat{p}_i - F(\underline{x}_i'\underline{\beta}), \end{aligned} \tag{2.17}$$

where $\hat{p}_i$ denotes the observed proportion of success for observation $i$. The posterior distribution of $\underline{\beta}$ determines the posterior distribution of residuals $r_i$. The posterior distribution of $r_i$ is not known analytically. However we can obtain the empirical posterior distribution of $r_i$ from the posterior distribution of $\underline{\beta}$. That is, let $\underline{\beta}^{(j)}, j = 1, 2, \ldots, m$ be the sample values, ${r_i}^{(j)}$ is computed according to

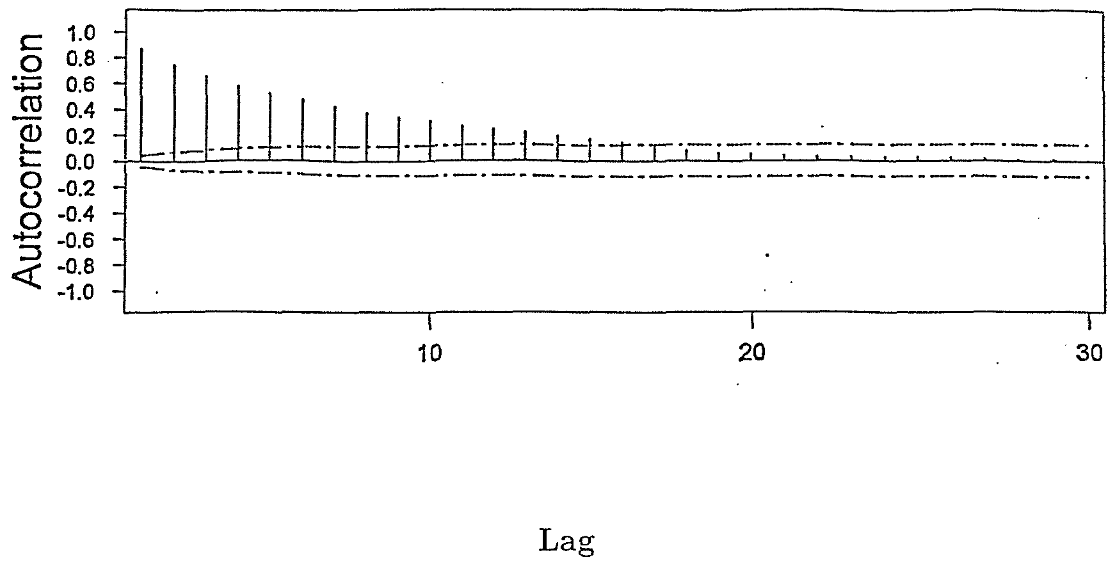
$$ {r_i}^{(j)} = \frac{y_i}{n_i} - F(\underline{x}_i' \underline{\beta}^{(j)}), \quad j = 1, 2, \ldots, m. $$

If the mean values of ${r_i}^{(j)}$ are close to $+1$ or $-1$, this indicates that the fit is not good. This method also works when $n = 1$ where ${r_i}^{(j)} = y_i - p^{(j)}$. However this method has a drawback that the sampling distribution is not known exactly. It is difficult to rely on the Bayesian residuals to account for the sample variation of the data. To determine whether a residual is actually extreme or not can be difficult. The method of latent residuals provide a good alternative way of diagnostic checking of the model.

## 2.5 Drawbacks of Adaptive Rejection Method

Despite the usefulness of the ARS method, it has two major drawbacks. The first drawback is that this method generates random variate only one at a time. Therefore for large models the computational speed is relatively slow. The second drawback is that for each output series, the serial correlation is rather high. That is the reason we use a sample size of 10,000. The following figure shows the sample autocorrelation functions up to lag 30 of parameter $\beta_0$ used in simulation studies. These autocorrelations are calculated based on the first 2000 generated values. The sample autocorrelation functions of $\beta_1$, $\beta_2$, and $\beta_3$ exhibited similar patterns of serial correlation.

Figure 2.8: Sample autocorrelation function of $\beta_0$




## 2.6 Conclusions

In conclusion, our proposed methodology is sound and viable. It is easy to implement if the marginal posterior distribution is log-concave. Zeger and Karim (1991) used rejection sampling method for the estimation procedure. Their link function is also the logit link. In fact the most popular link function for binary data is the logit as the model parameters may be interpreted as conditional log-odds ratios. As stated in their paper, generating coefficients of the random effects is the most time consuming step. However, the random effects coefficients can be directly generated using the ARS method. This is the most difficult part in the estimation. We use the ARS method in the Gibbs sampler scheme to generate the model parameters. The link functions are logistic. Albert and Chib (1996) used a probit link for analyzing binary longitudinal data. The probit link can also be generalized to multivariate responses ( Chib and Greenberg, 1998). Their paper provides a simulation-based and non-Bayesian analysis of correlated binary data using multivariate probit model. The posterior distribution is simulated by Markov chain Monte Carlo methods and maximum likelihood estimates are obtained by a Monte Carlo version of the EM algorithm. The Metropolis-Hastings algorithm is used in the Markov chain Monte Carlo methods. The probit link leads to simpler calculations, but may require many univariate normal random variate generations.

For practical purposes, probit and logistic regression curves look similar in many cases. It is not common to find examples, for which a logistic regression model fits well but the probit model fits poorly, or vise versa. Of course the parameter estimates differ for the two models, since their links have different scales. When both models fits well, the slope estimates in the logistic regression models are roughly about 1.6-2.0 times those in probit models. (see Agresti, 1996).

One further advantage of using ARS for Generalized linear mixed models is that this method can easily extend to deal with the general case where several mutually independent random effects $b_1, b_2, ..., b_c$ are incorporated simultaneously into the linear predictor $\eta = X\beta + Z_1 b_1 + \cdots + Z_c b_c$ where $b_l \sim N_{q_l}(0, \Sigma_l)$ $(l = 1, ..., c)$. The simplest but most commonly used covariance structure for each $\Sigma_l$ is of the form $\Sigma_l = \sigma_l^2 I_{q_l}$, where $\sigma_l^2 > 0$ is unknown and $I_{q_l}$ is a $q_l \times q_l$ identity matrix. This kind of model is known as GLMM *Model* I, GLMM *Model* II is where $\Sigma_l$ is unstructured. GLMM *Model II* has wide applications to statistical models for animal breeding experiments and biological sciences. Pan and Thompson (1998) used a quasi-Monte Carlo EM algorithm for parameter estimation based on maximum likelihood estimation (MLE). Their method is essentially numerical integration. Numerical integration becomes quite difficult if the dimension of the random component becomes very high. However, with a little more computational effort, our method can cope with this problem quite easily.

In the situation when model settings do not yield log-concave full conditionals, we cannot use the ARS method. However, Gilks *et al* (1995) have extended this method to deal with distributions that are not log-concave. Their method is called the adaptive rejection Metropolis sampling (ARMS) method. Software developed by the Medical Research Council Biostatistics Unit called "Bayesian inference Using Gibbs Scheme" (BUGS) can handle the simulation work in Section 2.3, but they do not use the Odell-Feivison (1966) technique to generate the variance-covariance matrix of the random effect component. Also BUGS does not automatically generate latent residuals. To analyze models with indicator variables such as the factor model for the BT data in the last section, it is more difficult to use the BUGS software.

# Chapter 3

# Random Effects Cumulative Logit Model

## 3.1 Introduction

In this chapter we extend the methodology used in last chapter to model ordinal response data. Ordinal response variables are very common in many fields of applications. A typical ordinal scale is labelled with words 'Good', 'Fair', 'Poor', or 'Bad'. In a psychology experiment this scale might be used to record different subjects' opinion of a mood stimulus. In medicine the measurement of interest might be a patient's reaction to a prescribed drug. In telecommunications research the ordinal response might be a subject's opinion of a telephone connection quality. Ordinal response data are often referred to as ordered categorical data.

An approach to modelling ordinal response data is to recognize explicitly that the responses are observations from a multinomial distribution. McCullagh (1980) proposed a family of models based on the cumulative probabilities of each category. McCullagh's approach is based on a generalized linear model. Generalized linear models with a cumulative link function are an excellent statistical apparatus to analyze the relationship between an ordinal response variable and the covariates.

For each subject $i$ we observe a response variable $Y_{ij}$ at the $j$th occasion, $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. $Y_{ij}$ may take on any one of $r$ ordered values labelled 1,2,...,r. A cumulative link model (Agresti, 1996) for these data would be of the form

$$\begin{aligned}\gamma_{ij,h} &= \Pr(Y_{ij} \leq h \mid \underline{x}_{ij},\ \underline{\beta}), \quad h = 1, 2, \ldots, r-1 \\ &= G(\alpha_h - \underline{x}'_{ij}\underline{\beta}) \qquad (3.1)\end{aligned}$$

where $G$ is the cumulative distribution function (CDF) of a continuous random variable taking values in $\Re$; $\alpha_h$, h=1,2,...,r-1, are ordered cut-points dividing the real line into r bins. $\underline{x}_{ij}$ is a vector of p explanatory variables for each $ij$th observation . $\underline{\beta}$ is a vector of coefficients of the covariates. If $G$ is the logistic cumulative distribution function, then the model is called a cumulative logit or proportional odds model (Agresti 1996).

If the ordinal response data are in the form of repeated measures on the same sampling units then the cumulative logit with random effects may be more appropriate. Hedeker and Gibbons (1994) have developed an appropriate methodology for the inclusion of random effects in cumulative link models. They propose a model for the continuous response underlying the repeated ordinal response. A similar approach is adopted in this thesis, but we regard the underlying continuous response as latent data. Hedeker and Gibbons (1994) assume that the distribution of the random effects is multivariate normal and the errors are assumed to be independent and normally distributed. This implies the cumulative probit model. If we use logit link then the errors assume a logistic distribution with mean 0 and variance $\sigma^2_\varepsilon$. Hedeker and Gibbons (1994) used maximum likelihood to estimate model parameters but this requires the evaluation of an integral as the random effect distribution needs to be integrated out of the likelihood. This integration is performed by Gaussian quadrature. Jansen (1990) applied the cumulative probit link model with random effects to an agricultural experiment. The data are clustered ordinal data. For parameter estimation Jansen (1990) uses Gaussian quadrature to perform the integration in the likelihood function. Ezzet and Whitehead (1991) used a random effects probit model to analyze longitudinal data from a cross-over trial. They also performed numerical integation to integrate the likelihood using FORTRAN.

Pan and Thompson (1998) use a Quasi-Monte Carlo EM algorithm to estimate model parameters in called generalized linear mixed models (GLMMs). GLMMs (Breslow & Clayton, 1993) are simply the usual generalized linear models with random effects. Booth and Hobert (1999) implement a what they called an " automated Monte Carlo EM algorithm" to maximize generalized linear mixed model likelihoods. Booth and Hobert (1999) claimed that their methods can be considerably more efficient than those based on Markov chain Monte Carlo algorithm. However, they also state that the methods proposed may break down when the intractable integrals in the likelihood function are of high dimension. In view of the difficulties in integrating out the likelihood when it comes to a very high dimensional problem, we propose that a Gibbs sampling approach provides a good alternative for parameter estimation. As in Chapter 2, the Gibbs sampler approach assumes a Bayesian formulation of the model.

## 3.2 Bayesian Approach

Albert and Chib (1993) present Bayesian implementations of the ordinal probit model. The link function $G$ is the normal cumulative distribution function $\Phi(\cdot)$. The Gibbs sampler is used to estimate the model parameters. Based on the work of Zeger and Karim (1991) and Albert and Chib (1993), we now present a Bayesian approach to a generalized model with random effects for ordered categorical data. Again the Gibbs sampler is used to estimate $\underline{\beta}$, the regression parameters and $\text{Var}(\underline{b}_i)$. In particular, we use the Adaptive Rejection Sampling (ARS) method introduced by Gilks and Wild (1992).

First of all, we rewrite equation 3.1 as

$$\gamma_{ij,h} \quad = \quad G(\alpha_h - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i) \tag{3.2}$$

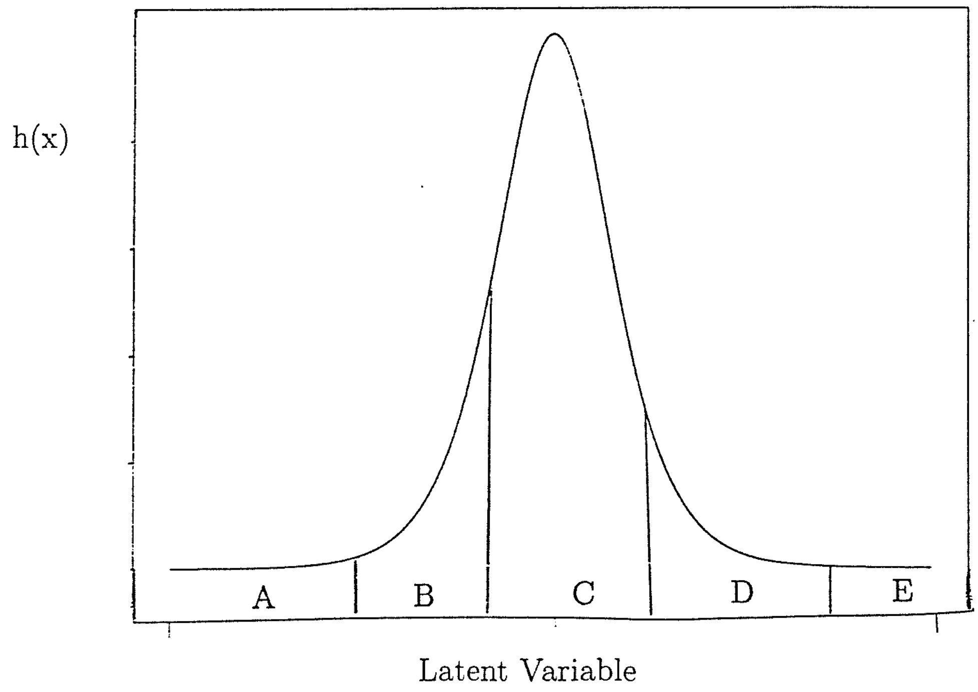
where $\underline{b}_i$ is the coefficient vector of the random component for each subject $i$ and $\underline{z}_{ij}$ is a subset of the explanatory variables $\underline{x}_{ij}$.

Then, $G$ is the logistic CDF

$$\text{logit}(\gamma_{ij,h}) \quad = \quad \alpha_h - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i \qquad (3.3)$$

Often $\alpha_h$ is referred to as the cut-point and one can assume that there exists a 'latent' (unobserved) continuous random variable $\lambda_{ij}$ such that $\lambda_{ij}$ follows certain continuous distribution with mean $\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$. We observe the ordinal response $Y_{ij}$ where $Y_{ij}$ = h if $\alpha_{h-1} < \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i \leq \alpha_h$. Also we define $\alpha_0$ = $-\infty$ and $\alpha_r$ = $+\infty$. Here, we assume $\lambda_{ij}$ follows the logistic distribution with mean $\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$ and standard variance. The graphical representation of the latent variable model is shown in Figure 3.1 below.

Figure 3.1: Graphical representation of the latent variable model




In Figure 3.1, the logistic distribution represents the distribution of a latent traits for particular individual. It is assumed that a random variable is drawn from this density, and the value of this random variable determines an individual's classification.

Now $\alpha_h, h = 1, 2, ..., r-1$, $\underline{\beta}$ and $\underline{b}_i$ are unknown parameters of our model. Assume that the random effects $\underline{b}_i$ follow a multivariate normal distribution with zero mean and unknown variance-covariance matrix $D$. We focus on the use of the Gibbs sampler to estimate all the unknown parameters.

## 3.3 The Model

For the Bayesian analysis of the model in equation 3.3, we take a diffuse prior for $(\underline{\beta}, \underline{\alpha}, \underline{b}_i)$. Albert and Chib (1993) fitted the Bayesian ordinal probit model. They noted that one cut-point had to be fixed and they choose $\alpha_1 = 0$. This is to ensure identifiability of the parameters. Nandram and Chen (1996) explain in detail the reason for fixing one cut-point. Without loss of generality, we also take $\alpha_1 = 0$ and the joint posterior density of $\underline{\beta}$ (vector of coefficient parameters of explanatory variables), $\underline{b}_i$ (vector of coefficients of random components of subject $i$ ), $\lambda_{ij}$ (continuous latent random variable) and $\underline{\alpha}$ (vector of cut-points) is then given by (up to proportionality)

$$\begin{aligned}\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij}) &= c \times P(\underline{\beta}, \underline{\alpha}, \underline{b}_i) \times \prod_{i=1}^{n} \prod_{j=1}^{k} \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h) \mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h) \right] \\ &\times \left[ \frac{e^{\lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i})^2} \right] \times |D|^{-\frac{1}{2}} \exp(-\frac{1}{2} \underline{b}_i' D^{-1} \underline{b}_i) \qquad (3.4)\end{aligned}$$

where $\underline{\beta}' = (\beta_0, \ \beta_1, \beta_2, \ ..., \beta_{p-1} \ )$, $\underline{\alpha}' = (\alpha_0, \ \alpha_1, \ \alpha_2, ..., \alpha_r \ )$, $\underline{b}_i' = (b_0, b_1, b_2, ..., b_{q-1})$ and $q < p$. $\mathrm{I}(\cdot)$ is the indicator function and $\lambda_{ij}$ is assumed to follow the logistic distribution with mean $\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$ and standard variance.

The conditional posterior distribution of $\underline{\beta}$, denoted as $[\underline{\beta} \mid \cdot]$, is then given by,

$$[\underline{\beta} \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij}) d\beta_0 d\beta_1 ... d\beta_{p-1}},$$

$$\begin{aligned}[\underline{\beta} \mid \cdot] &= c_1' \times P(\underline{\beta}, \underline{\alpha}, \underline{b}_i) \times \prod_{i=1}^{n} \prod_{j=1}^{k} \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h) \mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h) \right] \\ &\times \left[ \frac{e^{\lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i})^2} \right]\end{aligned}$$

where $c_1'$ is a constant and since the joint prior distribution is a diffuse prior; $\left[\underline{\beta} \mid \cdot\right]$ is given by

$$\left[\underline{\beta} \mid \cdot\right] \propto \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\frac{e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i})^2}\right] \tag{3.5}$$

i. e. $$\left[\underline{\beta} \mid D, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, y_{ij}\right] = \left[\underline{\beta} \mid \lambda_{ij}, \underline{b}_i, y_{ij}\right]$$

The conditional density of $\lambda_{ij}$, $[\lambda_{ij} \mid \underline{\beta}, D, \underline{\alpha}, \underline{b}_i, y_{ij}]$ , is given by

$$[\lambda_{ij} \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij}) d\lambda_{ij}},$$

$$\begin{aligned}\left[\lambda_{ij} \mid \underline{\beta}, D, \underline{\alpha}, \underline{b}_i, y_{ij}\right] &= c_2' \times P(\underline{\beta}, \underline{\alpha}, \underline{b}_i) \times \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h)\right] \\ &\quad \times \left[\frac{e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i})^2}\right], \end{aligned} \tag{3.6}$$

where $c_2'$ is also a constant and so we have

$$\begin{aligned}\left[\lambda_{ij} \mid \underline{\beta}, D, \underline{\alpha}, \underline{b}_i, y_{ij}\right] &\propto \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h)\right] \\ &\quad \times \left[\frac{e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i})^2}\right] \end{aligned} \tag{3.7}$$

i. e. $$\left[\lambda_{ij} \mid \underline{\beta}, D, \underline{\alpha}, \underline{b}_i, y_{ij}\right] = \left[\lambda_{ij} \mid \underline{\beta}, \underline{\alpha}, \underline{b}_i, y_{ij}\right]$$

The conditional posterior distribution of $\underline{b}_i$, $[\underline{b}_i \mid \cdot]$, is given by

$$[\underline{b}_i \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i \mid y_{ij}) db_1 db_2 ... db_q},$$

$$\left[\underline{b}_i \mid \underline{\beta}, D, \underline{\alpha}, \underline{b}_i, y_{ij}\right] \propto \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\frac{e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i})^2}\right] \times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1} \underline{b}_i) \tag{3.8}$$

$$[\underline{b}_i \mid \cdot] = \left[\underline{b}_i \mid \underline{\beta}, D, \lambda_{ij}, y_{ij}\right]$$

Further it is easily seen that

$$\left[D \mid \underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i\,, y_{ij}\right] = [D \mid \underline{b}_i]$$

Finally, the conditional density of $\alpha_h$ (the cut-point) given $\lambda_{ij}$, $\underline{\beta}$, $\underline{b}_i, D, y_{ij}$ and $\alpha_v$, $\nu$ $\neq$h, is given by Albert and Chib (1993) (up to proportionality ),

$$\begin{aligned}[\alpha_h \mid \cdot] \quad & \propto \prod_{i=1}^{n} \prod_{j=1}^{k} [\mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h) \\ & + [\mathrm{I}(y_{ij} = h+1)\mathrm{I}(\alpha_h \leq \lambda_{ij} < \alpha_{h+1})], \qquad (3.9)\end{aligned}$$

where $c_3'$ is also a constant. This conditional distribution can be seen to be uniform, i.e.

$$\begin{aligned}\alpha_h \sim \quad & \mathrm{Unif}\,[\max\{\max\{\lambda_{ij} : y_{ij} = h\}, \alpha_{h-1}\}, \\ & \min\{\min\{\lambda_{ij} : y_{ij} = h+1\}, \alpha_{h+1}\}] \qquad (3.10)\end{aligned}$$

After working out the necessary conditional densities, it is easy to draw up the Gibbs sampler scheme.

### 3.3.1 Log-concavity Conditions

To use the ARS method, it is necessary to check if the conditional posterior distributions satisfy the log-concavity condition. That is, for any density, $h$ (say), we need to check whether $\frac{\partial^2 \ln h}{\partial {\beta_v}^2} < 0$, where $\beta_v$ is the $v$th parameter of interest. First of all, consider the density of $\beta$, we let $\left[\underline{\beta} \mid \cdot\right] = h(\underline{\beta})$. Then

$$\begin{aligned}h(\underline{\beta}) \quad & = c_1' \times P(\underline{\beta}, \underline{\alpha}, \underline{b}_i) \times \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h)\right] \\ & \times \left[\frac{e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}}{(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i})^2}\right] \qquad (3.11)\end{aligned}$$

If we take natural logarithm on both sides of 3.11, then 3.11 becomes

$$\begin{aligned}\ln h(\underline{\beta}) \quad & = \ln c_1' + \sum_{i=1}^{n} \sum_{j=1}^{k} \ln \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \leq \alpha_h)\right] \\ & + \sum_{i=1}^{n} \sum_{j=1}^{k} (\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i) \\ & - 2\sum_{i=1}^{n} \sum_{j=1}^{k} \ln(1 + e^{\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i}), \qquad (3.12)\end{aligned}$$

and for fixed $g$, the first partial derivative becomes

$$\frac{\partial \ln h(\underline{\beta})}{\partial \beta_g} = -\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij,g} + 2\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij,g}\left[\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{(1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})}\right].$$

The second partial derivative is

$$\begin{aligned}\frac{\partial^2 \ln h(\underline{\beta})}{\partial \beta_g^2} &= -2\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij,g}^2\left[\left\{\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}\right\} - \left\{\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}\right\}^2\right] \\ &= -2\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij,g}^2(1-u_{ij})u_{ij} < 0, \qquad (3.13)\end{aligned}$$

where $u_{ij} = \frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}$ and it is easy to verify that $0 < u_{ij} < 1$.

$\lambda_{ij}$ can be directly sampled from a truncated logistic distribution.

To check the log-concavity condition of the coefficient of the random effects component, let's consider equation 3.8 and for fixed $i$, the natural logarithm of $h(\underline{b}_i)$ becomes

$$\begin{aligned}\ln h(\underline{b}_i) &= \ln c'_3 + \sum_{j=1}^{k}\ln\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \le \alpha_h)\right] \\ &+\sum_{j=1}^{k}(\lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i) - 2\sum_{j=1}^{k}\ln(1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}) \\ &-\frac{1}{2}\underline{b}_i{}'D^{-1}\underline{b}_i, \qquad (3.14)\end{aligned}$$

Taking the first partial derivative of $\ln h(\underline{b}_i)$ with respect to $b_u$, we have

$$\frac{\partial \ln h(\underline{b}_i)}{\partial b_u} = -\sum_{j=1}^{k} z_{ij,u} + 2\sum_{j=1}^{k} z_{ij,u}\left[\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}\right] - (b_u d_{uu} + \sum_{u\neq v} b_v d_{uv}),$$

where $d_{uu}$ and $d_{uv}$ are corresponding elements of $D^{-1}$.

$$\frac{\partial^2 \ln h(\underline{b}_i)}{\partial b_u^2} = -2\sum_{j=1}^{k} z_{ij,u}^2(1-u_{ij})u_{ij} - d_{uu} < 0; \qquad d_{uu} > 0. \qquad (3.15)$$

From the above results, the log-concavity condition is satisfied by each conditional posterior distribution concerned. The ARS method can therefore be used in the Gibbs sampler scheme. To sample Var($\underline{b}_i$), we use the Odell-Feivison (1966) technique.

In general, it can also be proved that $h(\underline{\beta})$ and $h(\underline{b}_i)$ are log-concave for any design matrix. The proof is stated as follows.

**Proof:**

Consider the log-conditional posterior distribution of $\underline{\beta}$, that is,

$$\begin{aligned}\ln h(\underline{\beta}) \quad &= \ln c_1' + \sum_{i=1}^{n}\sum_{j=1}^{k} \ln\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \le \alpha_h)\right] \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{k}(\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i) \\ &-2\sum_{i=1}^{n}\sum_{j=1}^{k}\ln(1+e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})\end{aligned}$$

To show this function is concave with respect to the coefficient vector $\underline{\beta}$, it suffices to show that $\ln(1+e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})$ is convex with respect to $\underline{\beta}$; since $\lambda_{ij} - \underline{x}_{ij}'\underline{\beta} - \underline{z}_{ij}'\underline{b}_i$ is a linear function with respect to $\underline{\beta}$ which is concave.

Now consider $\ln(1+e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})$, let

$$F_1(\underline{\beta}) = \ln(1+e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i}).$$

Then it can be proved that

$$\frac{\partial^2 F_1}{\partial\underline{\beta}\partial\underline{\beta}'} = u_{ij}(1-u_{ij})\mathbf{X}$$

where $u_{ij} = \left[\frac{e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i}}{(1+e^{\lambda_{ij}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})}\right]$ and $\mathbf{X} = \underline{x}_{ij}\underline{x}_{ij}'$ is a p×p symmetric matrix. It is easy to prove that for each $i$ and $j$, $0 < u_{ij} < 1$ and the symmetric matrix $\mathbf{X}$ is semi-positive definite. Therefore $F_1$ is convex. The calculus is the same as in the proof of Chapter 2.

Similarly, to show that the conditional posterior distribution of $\underline{b}_i$ is log-concave with respect to $\underline{b}_i$, let's consider equation 3.14, that is,

$$\ln h(\underline{b}_i) \quad = \ln c_3' + \sum_{j=1}^{k}\ln\left[\sum_{h=1}^{r}\mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1} < \lambda_{ij} \le \alpha_h)\right]$$

$$+\sum_{j=1}^{k}(\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i)-2\sum_{j=1}^{k}\ln(e^{1+\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})$$
$$-\frac{1}{2}\underline{b}_i{}'D^{-1}\underline{b}_i,$$

Let $F_2(\underline{b}_i)=2\ln(1+e^{(\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i)})+\frac{1}{2}\underline{b}'_iD^{-1}\underline{b}_i,\ j=1,2,...k$. To show that $\ln h(\underline{b}_i)$ is concave with respect to $\underline{b}_i$, it suffices to show that $F_2$ is convex. Along the above lines, it can be shown that

$$\frac{\partial^2F_2}{\partial\underline{b}_i\partial\underline{b}'_i}=2(1-u_{ij})u_{ij}\mathbf{Z}+D^{-1}.$$

where $\mathbf{Z}=\underline{z}_{ij}\underline{z}'_{ij}$ is a q×q symmetric matrix. The symmetric matrix $\mathbf{Z}$ and the inverse of variance-covariance matrix $D$ are semi-positive definite. Therefore $F_2$ is also convex. This completes the proof of log-concavity condition for any design matrix for the model.

## 3.4 Model Formulation Using Cumulative Probability

There is another method of modelling ordered categorical data. In section 3 of this chapter, a full-likelihood approach is being used. However, if we reconsider the joint posterior density of equation 3.4, that is,

$$\Pi(\underline{\beta},\underline{\lambda},\underline{b}_i\mid y_{ij}) = k\times P(\underline{\beta},\underline{\lambda},\underline{b}_i)\times\prod_{i=1}^{n}\prod_{j=1}^{k}\{\sum_{h=1}^{r}\{\mathrm{I}(y_{ij}=h)\,[\gamma_{ij,h}-\gamma_{ij,h-1}]\}$$
$$\times|D|^{-\frac{1}{2}}\exp(-\frac{1}{2}\underline{b}_i{}'D^{-1}\underline{b}_i)\} \qquad (3.16)$$

where $k$ is a constant and $\gamma_{ij,h}$ is the cumulative probability as defined in Section 3.2. Then

$$\Pi(\underline{\beta},\underline{\lambda},\underline{b}_i\mid y_{ij}) = k_1\times\prod_{i=1}^{n}\prod_{j=1}^{k}\{\sum_{h=1}^{r}\mathrm{I}(y_{ij}=h)$$
$$\left[\int_{-\infty}^{\alpha_h}\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{(1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})^2}d\lambda'_{ij}-\int_{-\infty}^{\alpha_{h-1}}\frac{e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{(1+e^{\lambda_{ij}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})^2}d\lambda'_{ij}\right]$$

$$\times |D|^{-\frac{1}{2}} \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i)$$

$$= k_1 \times \prod_{i=1}^{n}\prod_{j=1}^{k}\Big[(1+e^{\alpha_{h-1}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})^{-1} - (1+e^{\alpha_h-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})^{-1}\Big]^{I(y_{ij}=h)}$$

$$\times |D|^{-\frac{1}{2}} \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i), \qquad (3.17)$$

where $k_1 = k \times P(\underline{\beta}, \underline{\lambda}, \underline{b}_i)$ is a constant. In this way, we can see that $\lambda_{ij}'$, the so-called latent variable, vanishes after integration in this case. Therefore we do not need to estimate the unknown latent observations. We now have the following conditional posterior distributions:

$$(1) \quad [\underline{\beta} \mid \cdot] \equiv [\underline{\beta} \mid \underline{\alpha}, \underline{b}_i, y]$$

$$\propto \prod_{i=1}^{n}\prod_{j=1}^{k}\prod_{h=1}^{r}\left[\frac{1}{(1+e^{\alpha_{h-1}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})} - \frac{1}{(1+e^{\alpha_h-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})}\right]^{I(y_{ij}=h)},$$

$$(2) \quad [\underline{\alpha} \mid \cdot] \equiv [\underline{\alpha} \mid \underline{\beta}, \underline{b}_i, y]$$

$$\propto \prod_{i=1}^{n}\prod_{j=1}^{k}\prod_{h=1}^{r}\left[\frac{1}{(1+e^{\alpha_{h-1}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})} - \frac{1}{(1+e^{\alpha_h-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})}\right]^{I(y_{ij}=h)},$$

$$(3) \quad [\underline{b}_i \mid \cdot] \equiv [\underline{b}_i \mid \underline{\beta}, \underline{\alpha}, D, y]$$

$$\propto \prod_{j=1}^{k}\prod_{h=1}^{r}\left[\frac{1}{(1+e^{\alpha_{h-1}-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})} - \frac{1}{(1+e^{\alpha_h-\underline{x}_{ij}'\underline{\beta}-\underline{z}_{ij}'\underline{b}_i})}\right]^{I(y_{ij}=h)}$$

$$\times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i),$$

$$(4) \quad [D \mid \cdot] \equiv [D \mid \underline{b}_i]$$

$$\propto \prod_{i=1}^{n} |D|^{-\frac{1}{2}} \exp(-\tfrac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i).$$

### 3.4.1 Checking of Log-concavity Conditions for the Alternative Approach

In this subsection we need to show if all the conditional posterior densities satisfy the log-concavity conditions (up to proportionality) using the approach of cumulative probability. First of all, starting from the marginal conditional posterior distribution of $\underline{\beta}$, we let $h(\underline{\beta}) = \left[\underline{\beta} \mid \cdot\right]$ and taking the natural logarithm we have

$$
\begin{aligned}
\ln h(\underline{\beta}) &= k' + \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h) \ln\left[\frac{1}{(1+e^{\alpha_{h-1}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})} - \frac{1}{(1+e^{\alpha_{h}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})}\right] \\
&= k' + \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h) \ln\left[\frac{e^{\alpha_{h-1}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i} - e^{\alpha_{h}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{(1+e^{\alpha_{h}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})(1+e^{\alpha_{h-1}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i})}\right]
\end{aligned}
$$

where $k'$ is a constant.

For any $u$ and any fixed category $h$, we can show that

$$
\frac{\partial^2 \ln h(\underline{\beta})}{\partial \beta_u^2} = -\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij,u}^2 \left[w_1(1-w_1) + v_1(1-v_1)\right] \tag{3.18}
$$

where $w_1 = \frac{e^{\alpha_{h-1}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\alpha_{h-1}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}$ and $v_1 = \frac{e^{\alpha_{h}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}{1+e^{\alpha_{h}-\underline{x}'_{ij}\underline{\beta}-\underline{z}'_{ij}\underline{b}_i}}$. It is easy to prove that $0 < w_1 < 1$ and $0 < v_1 < 1$. The proofs for $h(\underline{b}_i)$ and $h(\alpha_h)$ are similar.

## 3.5 Adaptive Rejection Sampling (ARS)

Again the adaptive rejection sampling method (Gilks and Wild, 1992) is used for parameter estimation . For our model, the log-concavity conditions are satisfied in the full-likelihood approach as well as in the alternative approach using cumulative probability. Therefore, to estimate model parameters using the Gibbs sampler, either method can be used. Here we only report the results of augmented-likelihood approach. Alternatively, for the second method we could employ the Quasi-Monte Carlo EM algorithm (Pan and Thompson, 1998) to estimate the regression coefficient vector $\underline{\beta}$ and gauss-quadrature numerical integration method to estimate the random effects component.

Cowles (1996) and Nandram and Chen (1996) suggested modified methods in improving the convergence rate of generating the cut-point distribution in the case of ordinal probit model. Cowles (1996) presented a multivariate Hastings-within-Gibbs update step when generating latent data and bin boundary parameters jointly, instead of individually from their respective full conditionals. Cowles claimed that to generate the latent variables and the cut-points, her algorithm substantially improves Gibbs sampler convergence for large data sets. In Cowles' algorithm, the model parameters are partitioned into two sets. The latent variables and the cut-points are in one set and the regression coefficient vector $\underline{\beta}$ is in the other. Gibbs sampling is used to sample $\underline{\beta}$ and a Metropolis-Hastings sampling is then used to simulate the cut-points given $y$ (the observations) and $\underline{\beta}$. Finally the latent variable is sampled also by using Metropolis-Hasting method given $y, \underline{\beta}$ and the cut-points previously generated. The cut-points and latent variable are generated according to truncated normal densities over certain intervals. One may refer to the article by Cowles (1996) for more details. Nandram and Chen (1996) used Cowles' method, but reparameterize the cumulative-link generalized linear model to accelerate the convergence of Cowles' algorithm. They reparameterized the parameters by multiplying $\underline{\beta}$ and the latent variable by the reciprocal of the second cut-point. Thereby the new $\underline{\beta}$ and new latent variable follow different conditional distributions. They claimed one important advantage is that for the three bins problem it does not require the Hastings algorithm. Empirical results in their article show that their method improves Cowles' algorithm.

Here by implementing the ARS scheme, we found that through simulation studies, the ARS method captured fairly quickly the true parameters of $\underline{\beta}$ and once $\underline{\beta}$ converges, the cut-point distribution converges rapidly. The problem of slow mixing due to large number of categories and thus extreme narrow widths between cut-points does not happen in our situation. Convergence is not a problem in the ARS scheme for generating $\underline{\beta}$, the random effects component and the cut-points. In each iteration, we obtain updated values for these components. Therefore the problem of slow-mixing does not occur in the ARS scheme.

## 3.6 Simulation Studies

For simulation, we consider the following cumulative logit model with random effects:

$$\text{logit}\,\Pr(Y_{ij} \leq h \mid \underline{b}_i) = \alpha_h - (\beta_0 + \beta_1 t + \beta_2 x_i + \beta_3 t x_i + b_{0i} + b_{1i} t) \tag{3.19}$$

where $x_i = 0$ for half of the population and 1 for the remainder and t = -3, -2, -1, 0, 1, 2, 3. The fixed effects coefficients were set at $\beta_0 = +3, \beta_1 = -2.5, \beta_2 = +2.5$ and $\beta_3 = -1.0$. The random effects distribution is simulated with

$$\text{Var}(\underline{b}_i) = D = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.001 \end{bmatrix} \tag{3.20}$$

Ordinal response data are simulated according to equation 3.19 with $h = 1, 2, 3$ and cutpoints $= 0.0, 2.5, 5.0$. That is, this model has three cut-points and four ordered categories. Each data set comprised of n = 100 (subjects) clusters of size j = 7 for each $i$th subject. To perform Gibbs sampling, we run the chain for each parameter of interest for 11,000 times and discard the initial 1000 values as burn-in. Table 3.1 shows the overall results of a typical run using one simulated data set. The results in Table 3.1 indicate that the Gibbs sampler scheme gives reasonable inferences. The overall results are quite good. Only $\beta_1$ has slight negative bias and $\text{Var}(b_0)$ is a slight underestimate.

Table 3.1. Results of Simulation Studies

| Parameter | True value | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| $\beta_0$ | 3.0 | 3.0781 | 3.0923 | 0.2753 | 2.3109 | 4.8921 |
| $\beta_1$ | -2.50 | -2.6912 | -2.7081 | 0.2017 | -3.2910 | -2.2451 |
| $\beta_2$ | +2.50 | 2.5807 | 2.4591 | 0.7906 | 1.1702 | 7.2156 |
| $\beta_3$ | -1.00 | -1.0152 | -0.9618 | 0.4378 | -3.0871 | 0.0714 |
| Var($b_0$) | +1.00 | 0.7231 | 0.5891 | 0.5723 | 0.0917 | 6.7619 |
| Cut2 | 2.5 | 2.3971 | 2.4201 | 0.2245 | 1.9092 | 2.9377 |
| Cut3 | 5.0 | 4.6658 | 4.7134 | 0.2835 | 4.0280 | 5.2948 |

To monitor convergence, we use the method of batching for each individual series and we also calculated the potential scale reduction factor (PSRF) R. In this simulation case R=1.0083 which indicates that the Gibbs sampler scheme converged very well. Since this is a simulation study and the value of R is close to 1, we have not reported figures for batching.

## 3.7 Analysis of Telephone Connection Quality

In Section 2.6 of last chapter we discussed the BT experimental data. In this chapter we analyze the ordinal response of the experiment concerning telephone connection quality. A frequently used method of response in British Telecom experiments on transmission assessment is a five -point scale, graded subjectively from 'Excellent' to 'Bad'. This and very similar variations are the recommended scales of use in telecommunications work. The traditional British Telecom method of analysis for opinion score responses has been to perform an analysis of variance on the numerical scores assigned to the categories (0 to 4). From a practical view this has been done with success. One of the assumptions underlying the analysis-of-variance procedure is that the response variable follows a Normal distribution. The opinion score is constrained to one of five values, i.e., it is not a continuous response but rather a discrete ordered one. Approximating a discrete response with a five values by a Normal curve is a rather crude approach. Also the scores attributed to different categories are arbitrary. These two considerations have the consequence of inefficient estimation of parameters by the standard analysis-of-variance approach.

In this Chapter, we analyze the E198 experiment by fitting the cumulative logit model with random effects based on a Bayesian approach. The E198 experiment has limited duration of conversation time for each subject. Wolfe (1996) has analyzed this data set by fitting a random effects cumulative model. His approach is basically a frequentist approach. In his model, all the cut-points are directly estimated, but his model does not include the intercept term. In our latent variable model, the first cut-point has to be fixed at zero for model identifiability. The latent variables $\lambda_{ij}$ are generated from the logistic distribution with mean $\underline{x}_i'\underline{\beta}-\alpha_1$ and standard variance. An intercept term is included in our model. However the two models are equivalent. Only the methods of estimation are different. Therefore if we rescale our final estimates of the latent variable model, we should arrive at 'approximately' the same estimates that Wolfe (1996) has obtained in his Thesis. One has to remember that there are Monte Carlo errors in the Gibbs sampling scheme. In the following we present the final results of our analysis. The model setting is the same as in Wolfe (1996). As in the simulation study, we generated 10,000 random variates for each parameter after 1000 burn-in values. In Table 3.2 we show the results of our estimates for the below.

Table 3.2. Results of random effects cumulative logit model for the BT (E198) experiment

| para. | mean | s. d. | s.d.[1] | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|---|
| Intercept | 8.2943 | 0.6057 | 0.2872 | 7.5728 | 8.9803 |
| Cond 1 | aliased | - | - | - | - |
| Cond 2 | -0.0280 | 0.7190 | 0.0181 | -0.9776 | 0.9144 |
| Cond 3 | -0.3606 | 0.7186 | 0.0145 | -1.3362 | 0.6664 |
| Cond 4 | -6.8385 | 0.7159 | 0.0209 | -7.7756 | -5.8589 |
| Cond 5 | -0.5684 | 0.7407 | 0.0178 | -1.6063 | 0.4543 |
| Cond 6 | -2.6649 | 0.6991 | 0.0167 | -3.5526 | -1.6952 |
| Cond 7 | -6.1285 | 0.7074 | 0.0134 | -7.0683 | -5.1584 |
| Cond 8 | -7.2762 | 0.6959 | 0.0161 | -8.2016 | -6.3235 |
| R.E. Var | 1.5992 | 0.5568 | 0.0534 | 1.1200 | 2.1772 |
| Cut 1 | set | at | 0 | | |
| Cut 2 | 3.8102 | 0.7209 | 0.0872 | 2.9375 | 5.2712 |
| Cut 3 | 6.6874 | 0.69087 | 0.0178 | 5.8420 | 7.9656 |
| Cut 4 | 9.1315 | 0.7192 | 0.0461 | 7.9064 | 10.4129 |

1:This is the standard deviation of batching means

The results in Table 3.2 can be compared with those obtained by Wolfe (1996). He used GLIM4 software to carry-out the fittings. The standard errors of the Monte Carlo results are generally smaller than those obtained from GLIM4 which used weighted least square method. The following table shows Wolfe's results

Table 3.3. Results of Wolfe's (1996) estimates (E198) experiment*

| Parameter. | Parameter estimate | Standard error |
|---|---|---|
| Cut-point 1 | 0 | - |
| Cut-point 2 | 3.31 | - |
| Cut-point 3 | 6.11 | - |
| Cut-point 4 | 8.75 | - |
| Cond 1 | 0 | - |
| Cond 2 | -0.36 | 0.75 |
| Cond 3 | -0.58 | 0.59 |
| Cond 4 | -6.67 | 1.07 |
| Cond 5 | -0.33 | 0.90 |
| Cond 6 | -2.68 | 0.81 |
| Cond 7 | -5.89 | 0.94 |
| Cond 8 | -8.45 | 0.98 |
| R.E. Var | 1.11 | 0.34 |

*: This table is taken from Table E.2 in Appendix E of the Wolfe's (1996) thesis.

The results obtained by Markov chain Monte Carlo method are similar to those obtained by Wolfe (1996). As we can see from Table 3.2, the batching variances are very small relatively to overall sample variances in each chain. The number of sample points in each batch is 500. There are 20 batches. The PSRF R=1.0178 for Table 3.2. This indicates that that convergences are good in each of the Gibbs sampling scheme. The last two columns in each of Table 3.2 show the 2.5% and 97.5% quantile values. In the following, we present the empirical posterior distribution of the intercept, condition 2 effect, condition 8 effect and random effects variance.

Figure 3.2: Empirical Posterior

Distribution of Intercept

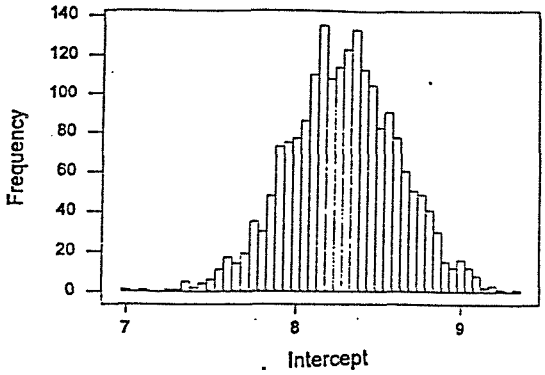



Figure 3.3: Empirical Posterior

Distribution of Condition 2

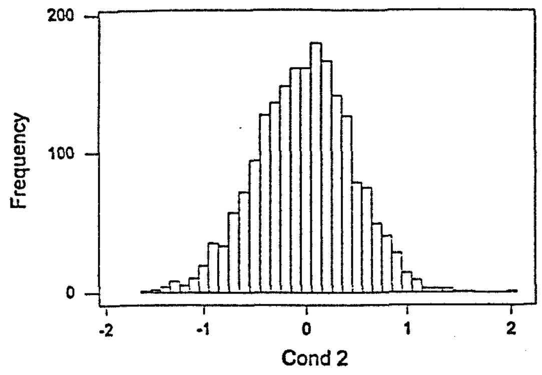



Figure 3.4: Empirical Posterior

Distribution of Condition 8

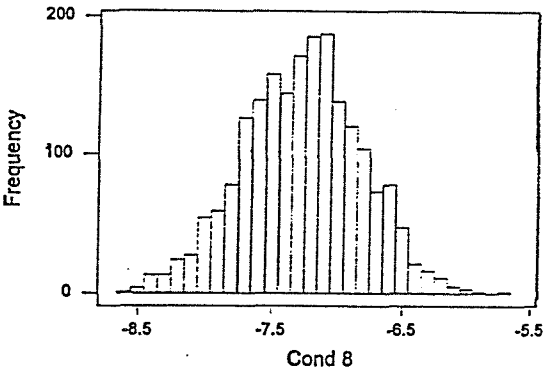
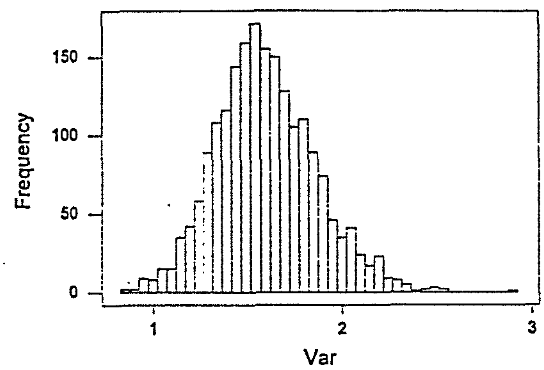



Figure 3.5: Empirical Posterior

Distribution of Random Effect Variance




Also the scatter plots of random effect variance against the intercept and condition 8 effects are shown on Figures 3.6 and 3.7 respectively. These two graphs show the correlation between the random effect variance and mean location level and condition 8 is negligible.

Figure 3.6: Scatter Plot of

Var($b_0$) vs Intercept

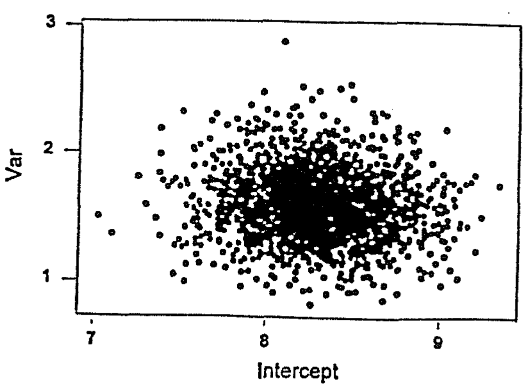
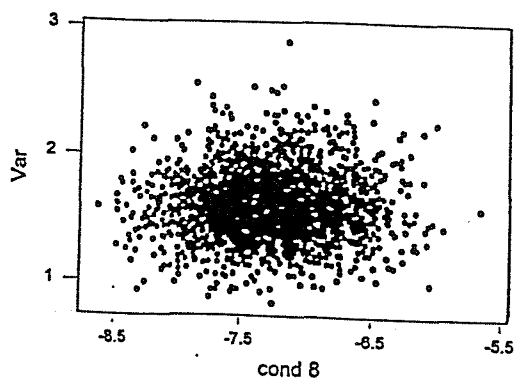



Figure 3.7: Scatter Plot of

Var($b_0$) vs Cond. 8



### 3.7.1 Posterior-Predictive Probability Distributions

As we have demonstrated in Chapter 2 we can compute the predictive probabilities for future events from the posterior-predictive density. For the cumulative logit model of E198 experiment, we compute the posterior-predictive distribution of scoring a bad telephone connection for subject 8 and subject 16 at each condition level. These are important figures because test-engineers would like to know which transmission condition is likely to cause a bad connection and the associated probabilities. There are eight conditions. At each condition $k$, the observation is binary ($n_k = 1$ for each k). The following Tables show the posterior- predictive distribution of subject 8 and subject 16. These figures are calculated based on the simulated values of $\underline{\beta}^{(j)}, j = 1, 2, \ldots, m$. Sample size $m$ is also equal to 2,000. The sample is taken from the last 2,000 values of the entire simulated Markov chain.

Table 3.4. Predictive probabilities of having a bad telephone connection of subject 8 at each given conditions (E198 experiment).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s.d. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0147 | 0.1113 | 0.0024 | 0..0484 |
| Cond 2 | 0.0010 | 0.0344 | 0.0001 | 0.0041 |
| Cond 3 | 0.0013 | 0.0341 | 0.0001 | 0.0041 |
| Cond 4 | 0.4756 | 0.4472 | 0.1158 | 0.8337 |
| Cond 5 | 0.0011 | 0.0342 | 0.0001 | 0.0044 |
| Cond 6 | 0.0261 | 0.1588 | 0.0029 | 0.0994 |
| Cond 7 | 0.2697 | 0.4052 | 0.0443 | 0.6549 |
| Cond 8 | 0.5404 | 0.4431 | 0.1604 | 0.8711 |

Table 3.5. Predictive probabilities of having a bad telephone connection of subject 16 at each given conditions (E198 experiment).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0116 | 0.1085 | 0.0012 | 0.0455 |
| Cond 2 | 0.0008 | 0.0333 | 0.0001 | 0.0035 |
| Cond 3 | 0.0009 | 0.0326 | 0.0001 | 0.0036 |
| Cond 4 | 0.4093 | 0.4532 | 0.0696 | 0.8228 |
| Cond 5 | 0.0009 | 0.0372 | 0.0001 | 0.0043 |
| Cond 6 | 0.0206 | 0.1513 | 0.0017 | 0.0804 |
| Cond 7 | 0.2226 | 0.4006 | 0.0275 | 0.6309 |
| Cond 8 | 0.4709 | 0.4599 | 0.0979 | 0.8709 |

Results in Table 3.4 and 3.5 both show that condition 4 and 8 are the most likely conditions that will give a poor connection. While condition 7 has about 20% chance of causing a bad line, the rest of the transmission conditions have very low chance of getting a bad connection.

### 3.7.2 Latent Residuals Analysis

Finally for model checking, we use the method of latent residuals introduced by Albert and Chib (1995). Their method is applied for models for a binary variable only; but it is not difficult to extend their method to models for ordinal categorical variable. That is if equation 3.3 is the correct model for the data,

$$\lambda_{ij} = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i + \varepsilon_{ij} \tag{3.21}$$

and $\lambda_{ij}$ are the latent variables for the h category, $\alpha_{h-1} < \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i < \alpha_h$ where $\alpha_{h-1}$ and $\alpha_h$ are the corresponding cut-points. As mentioned in Section 3.2, the latent variable $\lambda_{ij}$ follows a logistic distribution with mean $\underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i$ and unit variance. $\lambda_{ij}$ can be generated directly from the ARS method for given values of $\underline{\beta}^{(t)}$

and $\underline{b}_i^{(t)}$ at the $tth$ stage of iteration of our Gibbs sampler scheme as we have done in Section 3.5. Therefore latent residuals $\varepsilon_{ij}$ are defined as

$$\varepsilon_{ij} = \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i, \tag{3.22}$$

Latent residuals $\varepsilon_{ij}$ are a priori a random sample from a standard logistic distribution. To show that the generated latent residuals of our model followed a standard logistic distribution, a Kolmogorov-Smirnov test is used to test whether $\varepsilon_{ij}$ do follow a standard logistic distribution. To conduct the test a random sample of size 30 is being picked from the 5000 generated latent residuals each time and we obtain the necessary $D_n$ test statistic as well as Prob$(D \leq D_n)$. The test is repeated for 50, 100 and 1000 times. Table 3.3 show the overall results for E198.

Table 3.6: Results of K-S Test of latent residuals
(E198 Experiment)

| Sample size (n) | No. of runs (m) | Mean $D_n$ | Mean $\Pr(D \leq D_n)$ |
|---|---|---|---|
| 30 | 50 | 0.2025 | 0.2127 |
| 30 | 100 | 0.2189 | 0.1650 |
| 30 | 1000 | 0.2091 | 0.1979 |

The histogram of mean residuals for each observation of each subject are shown in Figure 3.8 and in Figure 3.9 we show the mean latent residuals plot against each subject. The empirical distribution of the 256 mean latent residuals is superimposed by a standard logistic distribution curve. From looking at Figure 3.8, the empirical distribution is in close agreement with the underlying theoretical distribution. However if we inspect the mean residuals plot of Figure 3.9, there are quite a number of mean latent residuals fall outside the $\pm$ 3 limits. This suggests that a new model should be considered for improvement of model fitting. We will consider a new model for these data in next chapter 4.

Figure 3.8: Histogram of Mean Residuals

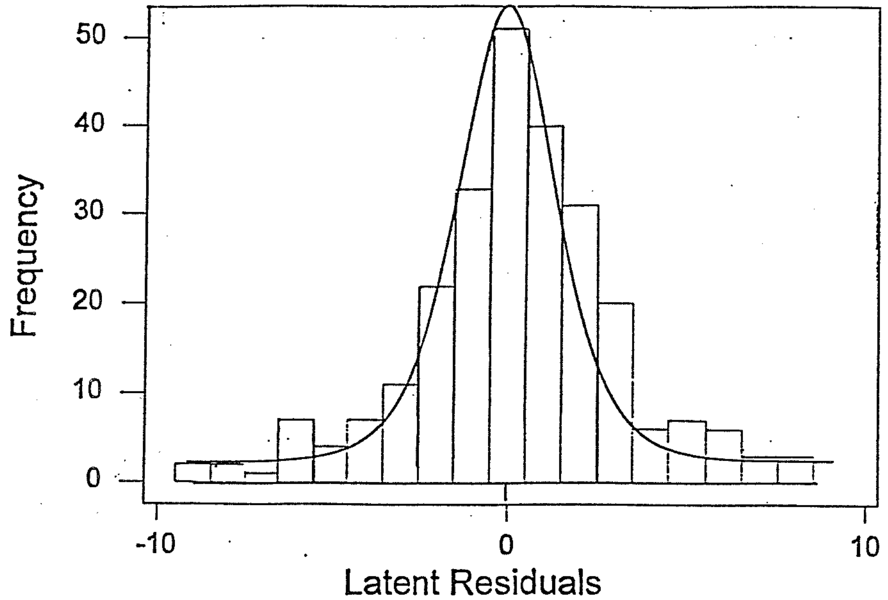
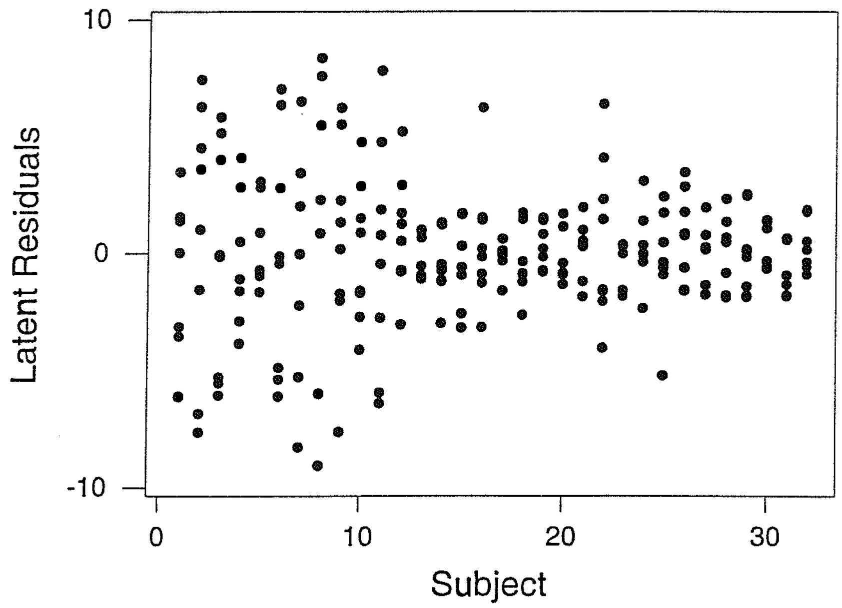



Figure 3.9: Plot of Mean Latent Residuals against Subject




## 3.8 Cumulative Rasch Model

It is well known that the cumulative Rasch model (Agresti and Lang, 1993) is a special case of the random effect cumulative logit model and the Rasch item response model (Rasch, 1960) is also a special case of the cumulative Rasch model. The Rasch item response model caters for repeated binary responses with subject-specific effects while the cumulative Rasch model caters for repeated ordered categorical responses.

To begin with the cumulative Rasch model, it is assumed that there are n subjects making $k$ repeated responses, $y_{ij} = h, i = 1, 2, ..., n$ and $j = 1, 2, ..., k$, on the same ordered categorical scale. For subject $i$ and response measure $j$, let $\pi_{ij,h}$ denote the probability of response in category $h$, for $h = 1, 2, ..., r$ and let $\gamma_{ij,h} = \pi_{ij,1} + \cdots + \pi_{ij,h}$. Now let's consider the cumulative logit model

$$\text{logit}(\gamma_{ij,h}) = \lambda_h + \alpha_i - \beta_j, \tag{3.23}$$

where $\lambda_h, h = 1, 2, ..., r-1$ are the 'cutpoints'. $\alpha_i$ and $\beta_j$ are respectively the subject and response effect parameters being independent of $h$. If the model holds, for each $i$ and $j$, there is an underlying continuous response that has a logistic distribution with mean $\beta_j - \alpha_i$ and the observed response falls in category $h$ when the underlying response falls between $\lambda_{h-1}$ and $\lambda_h$. It is necessary to set $\lambda_1 = \beta_1 = 0$ for identifiability of the model. Using the Bayesian formulation and the Gibbs sampler approach one can in fact estimate (i) the various cutpoints $\lambda_h$, $h = 2, ..., r-1$, (ii) $\beta_j$, $j = 2, ..., k$ and (iii) $\text{Var}(\alpha_i)$. Agresti and Lang (1993) used a maximum likelihood approach to estimate model parameters. The methodology that is employed for ordinal data in previous sections of this Chapter can be used for cumulative Rasch model.

### 3.8.1 The Rasch Model

The Rasch Model (Rasch, 1960) explains the occurrence of a data matrix containing the binary scored answers of a sample of n persons (the subjects) to a fixed set of $k$ items. This model has been widely used in social and behavioural sciences. In a psychological test or attitude scale, one tries to measure the extent to which a person possesses a certain property such as intelligence, arithmetic ability etc. Such properties are often called 'latent traits'. The use of a test or scale presupposes that one can indirectly infer a person's position on a latent trait from his/her responses to a set of well-chosen items. A statistical model of the measurement process should allow us to make predictions of future behaviour when confronted with other items from the same domain.

In a Rasch model a $n \times k$ data matrix is obtained containing the binary scored answers, i.e. $y_{ij} = 0$ or $1, i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. In other words, there are $S_i,\ i = 1, 2, ..., n$, subjects and a fixed set of $I_j,\ j = 1, 2, ..., k$ items. Each item $I_j$ has a real-valued item parameter $\beta_j$ denoting the difficulty of that item. Further let $\lambda_{ij}$ be denoted as the latent trait variable which assumed to follow a logistic distribution with mean $\beta_j - \alpha_i$ and standard variance. $\alpha_i$ is the subject-specific effect. Therefore if we assume a diffuse prior, the joint posterior distribution for the logit model

$$\text{logit}\,\Pr(y_{ij} = 1) = \lambda_{ij} + \alpha_i - \beta_j \tag{3.24}$$

is given by

$$\begin{aligned}\Pi(\underline{\beta}, \lambda_{ij}, \alpha_i \mid y_{ij}) &= c \times P(\underline{\beta}, \lambda_{ij}, \alpha_i) \times \prod_{i=1}^{n} \prod_{j=1}^{k} \left[ \sum_{h=0}^{1} I(y_{ij} = h) I(\lambda_{ij} \leq 0 \ \ or \ \ \lambda_{ij} > 0) \right] \\ &\times \frac{e^{\alpha_{ij} - \alpha_i - \beta_j}}{(1 + e^{\lambda_{ij} - \alpha_i - \beta_j})^2} \times \frac{1}{\sigma} e^{-\frac{1}{2} \frac{\alpha_i^2}{\sigma^2}} \end{aligned} \tag{3.25}$$

The mathematics for deriving the various conditional posterior distribution and log concavity conditions follows from the cumulative Rasch model. In summing up, we have also provided a methodology to model the cumulative Rasch model which is used to model ordered categorical data in psychology and behavioural science. This method can be used for the ordinary Rasch model for binary responses.

## 3.9 Conclusions

Our proposed methodology for dealing with ordinal data which are commonly found in many practical situations is practicable. It is easy to implement if the conditional posterior distributions are log-concave. Convergence seems not to be a stumbling obstacle in the analysis. Often convergence is achieved at a very fast rate despite the drawbacks we mentioned in chapter 2. In the situation when model settings do not yield log-concave full condition, we can use the adaptive rejection Metropolis sampling (ARMS) method also introduced by Gilks *el al.* (1995).

The well-known Rasch Model (Rasch, 1960) in Psychology is simply a special case of this model. Bayesian analysis with the Gibbs sampler provides a good alternative to the classical frequentist approach. Lastly, concerning convergence rate, our method is a more direct method than those proposed by Nandram and Chen (1996) and Cowles (1996); in terms of sampling from various conditional posterior distributions.

Furthermore, our method can easily be extended for the cumulative logit model with complicated random effect components. That is, it can fit models of the following form

$$\text{logit}(\gamma_{ij,h}) = \lambda_h - X\underline{\beta} - Z_1 b_1 \cdots - Z_c b_c, \qquad (3.26)$$

where $b_l \sim N_{q_l}(0, \Sigma_l)$ $(l = 1, ..., c)$ and $\Sigma_l$ is either an diagonal or unstructured matrix. This model is an extension of the GLMM *Model II* mentioned in chapter 2 for ordinal data. This is a research area which is worth further investigation. Again the BUGS software can handle the simulation work for the cumulative logit model, but again they do not use the Odell-Feivison (1966) technique to generate variance-covariance matrix of the random effect component. To analyze a model with indicator variables such as the factor model for BT data in last section, it will be a good attempt to use the BUGS software.

# Chapter 4

# Random Effects Cumulative Logit Model with Subject-Specific Scaling Term

## 4.1 Introduction

In the E198 experiment, subject-differences in the scores (responses) are quite significant. These phenomena are found firstly in the higher or lower scoring by the subjects on the response scale and secondly by clustering or spreading to the extremes of responses by the subjects. To show this phenomena, it is easiest to examine the histogram of the responses of those subjects who exhibit spreading in the extreme categories or clustering on one or two particular categories. For example subject 1, 2, 8 and 15 have a tendency of scoring the highest category (see histograms below).

Figure 4.1: Histogram of Scores (Sub. 1)

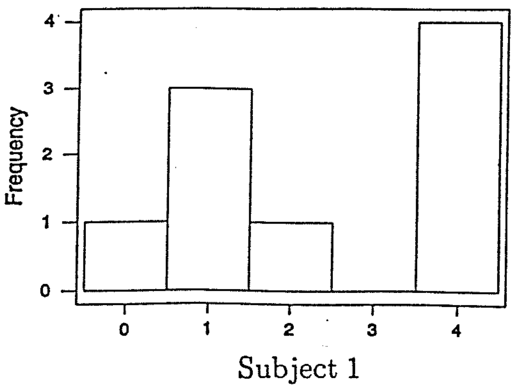



Figure 4.2: Histogram of Scores (Sub. 2)

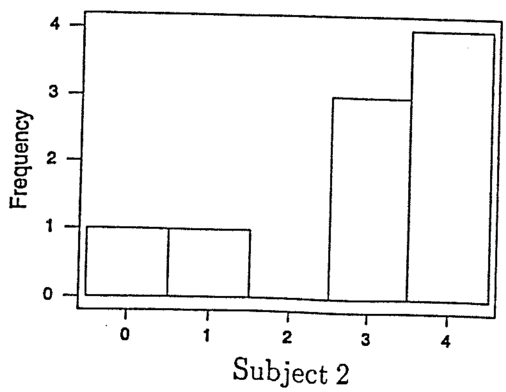


Figure 4.3: Histogram of Scores (Sub. 8)

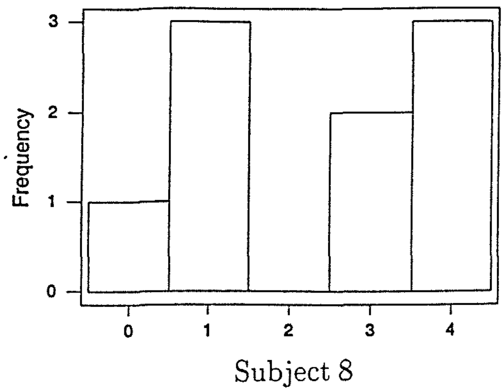
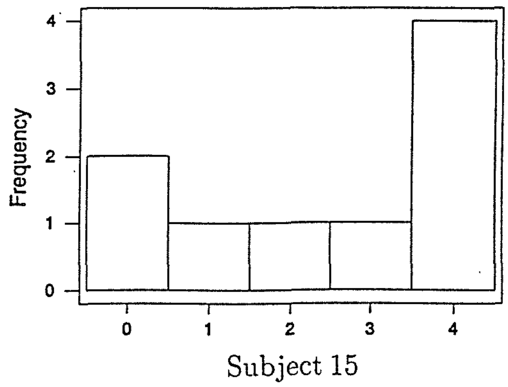



Figure 4.4: Histogram of Scores (Sub. 15)




On the other hand, subject 3, 7, 9, 12 and 13 have similar patterns (scoring more on category 3), but none has scored the highest category. Subject 14 also has not scored the highest category, but this subject is more likely to score the lowest category (see histograms below).

Figure 4.5: Histogram of Scores (Sub. 3)

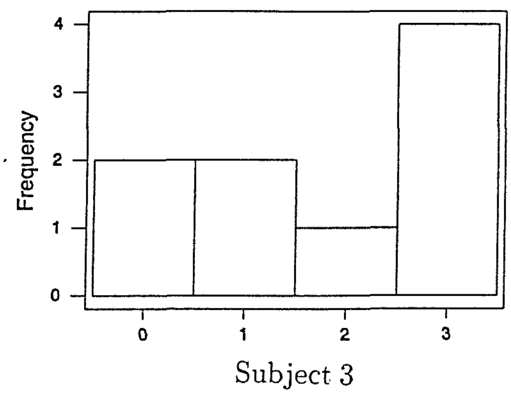



Figure 4.6: Histogram of Scores (Sub. 7)

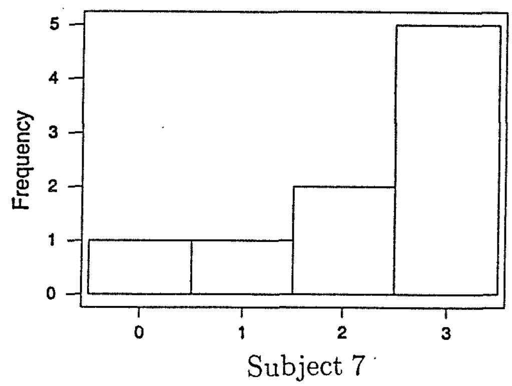



Figure 4.7: Histogram of Scores (Sub. 9)

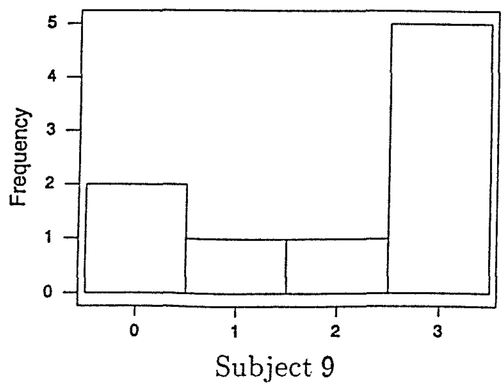



Figure 4.8: Histogram of Scores (Sub. 12)

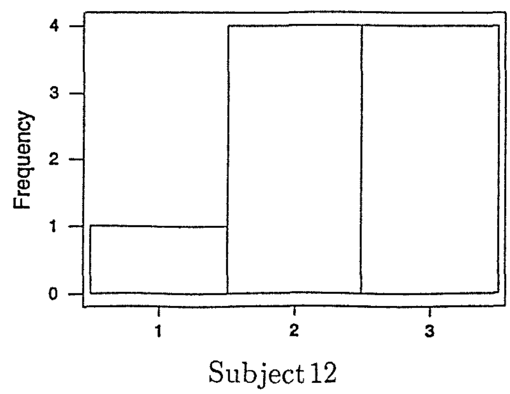


Figure 4.9: Histogram of Scores (Sub. 13)

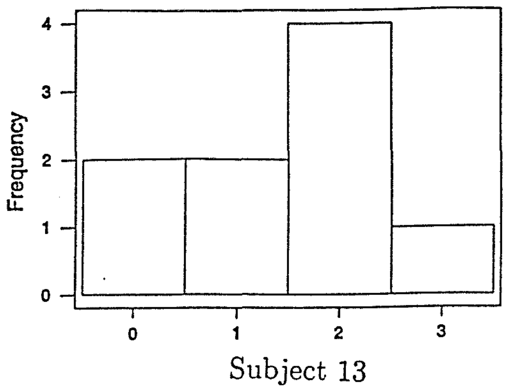
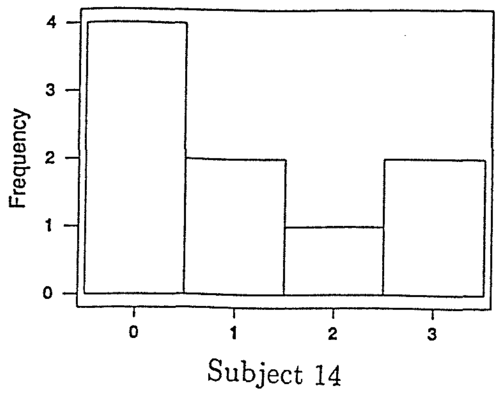



Figure 4.10: Histogram of Scores (Sub. 14)




Therefore we suspected that a model with a random scaling term should be used to explain this phenomenon as each response subject tends to have his/her own ordering scale in mind. There are two forms of model with a scaling term. These are

$$\log\left(\frac{\gamma_{ij,h}}{1-\gamma_{ij,h}}\right) = \frac{\alpha_h - \underline{x}_{ij}^T\underline{\beta}}{\exp(\tau_i)} \tag{4.1}$$

which is presented by McCullagh (1980). The quantity $\underline{x}_{ij}^T\underline{\beta}$ is called the "location" for the $i$th row and $\tau_i$ is called the "scale" for the $i$th row. In general, such a model is appropriate only when the number of response categories is at least three.

The other model is

$$\log\left(\frac{\gamma_{ij,h}}{1-\gamma_{ij,h}}\right) = \alpha_h \exp(\tau_i) - \underline{x}_{ij}^T\underline{\beta} \tag{4.2}$$

which is discussed by Kijewski *et al.* (1989). The scaling term quantifies how much responses are spread out (or concentrated) across the ordinal scale. Hence a subject-specific scaling term quantifies the type of difference between different subjects. The difference between models by McCullagh (1980) and Kijewski *et al.* (1989) can be illustrated by the following two examples.

Consider the McCullagh's (1980) model without random effects and imagine a set of data which can be divided into 2 groups according to the values of a binary characteristic. Further, consider a cumulative logit model with a scaling term

$$\text{logit}(\gamma_{1j,h}) = \frac{(\alpha_h - \beta_1)}{\exp(\tau_1)},$$

$$\text{logit}(\gamma_{2j,h}) = \frac{(\alpha_h - \beta_2)}{\exp(\tau_2)},$$

$$\beta_1 = 0 \qquad\qquad \tau_1 = 0.$$

The binary explanatory variable has both a location and scale parameter associated with it. The responses are split into 2 groups according to the value of the binary covariate. The responses in these 2 groups are assumed to be observations from 2 underlying distributions. The difference between the underlying distributions is described via the cumulative logit model in two ways. The first way is by the difference in their location, described by $\beta_2$ and the second way is by the difference in their dispersion, described by $\tau_2$.

Now consider the model discussed by Kijewski *et al.* (1989). Also a simple example of the cumulative logit model with only scaling term with continuous covariate is considered. That is

$$\text{logit}(\gamma_{1j,h}) = \alpha_h \exp(\tau_1) - \beta_1,$$

$$\text{logit}(\gamma_{2j,h}) = \alpha_h \exp(\tau_2) - \beta_2,$$

$$\text{with} \qquad \tau_1 = 0.$$

The $\exp(\tau_i)$ term has a multiplicative effect on the cut-point parameters $\alpha_h$. The value of $\tau_2$ has the effect of either clustering the cut-points on the underlying continuum (if $\tau_2 < 0$) or spreading them out (if $\tau_2 > 0$).

The inclusion of a scaling term in the cumulative logit model (Kijewski *et al.*; 1989) is more straightforward than in the model introduced by McCullagh (1980). This is because the cumulative logit model can be developed by assuming that the ordinal response is a manifestation of an underlying continuous response and that the threshold parameters indicate cut-points on the underlying continuum between categorical responses. Necessarily the threshold parameters are ordered in this case.

Wolfe (1996) proposed a cumulative logit model with form:

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \alpha_h \exp(\zeta_s \sigma_\zeta) - \omega_s \sigma_\omega - \mathbf{X}_i \beta$$

where $\omega_s$ is a random effect describing subject location, $\zeta_s$ is a random effect describing subject scaling and these two random effects are assumed to come from two independent standard normal distributions. The amount of random dispersion is measured by $\sigma_\zeta$ and $\sigma_\omega$. However Wolfe (1996) did not do any work on simulation or on practical data analysis. He wrote on page 115 of his thesis, " Unfortunately time ran out before this could be attempted as a contribution to this thesis. The development of general purpose software to fit the cumulative logit model with both random location and scale effects would be of practical use".

## 4.2 Bayesian Model Formulation

We now present a Bayesian approach to modelling ordinal data with a random effects cumulative logit model with random scaling term. This model is based on the form developed by Kijewski *et al.* (1989).

First of all, consider the cumulative probability, $\gamma_{ij,h}$, for the response categories less than or equal to h, i.e. Prob$(Y_{ij} \leq h)$

$$\gamma_{ij,h} \quad = \quad G(\alpha_h \exp(\tau_i) - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i) \qquad (4.3)$$

where $\tau_i$ is a scaling effect random variable and $\underline{b}_i$ is the coefficient vector of the random component for each subject $i$ and $\underline{z}_{ij}$ is a subset of the explanatory variables $\underline{x}_{ij}$. Furthermore

$$\begin{aligned} \text{logit}(\gamma_{ij,h}) &= \alpha_h \exp(\tau_i) - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i \\ &= \alpha'_h - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i, \quad \alpha'_h = \alpha_h \exp(\tau_i). \end{aligned}$$

$\alpha_h$ is referred to as the cut-point for the whole population and $\alpha'_h$ is the $i$th individual cut-point. It is already assumed that there exists a "latent" (unobserved) continuous random variable $\lambda'_{ij}$ such that $\lambda'_{ij}$ follows a logistic distribution with mean $\underline{x}_{ij}^T\underline{\beta}+\underline{z}_{ij}^T\underline{b}_i$ for each individual. For the scaling effect random variable $\tau_i$, we assume that $\tau_i$ follows a normal distribution with mean 0 and variance $\sigma_\tau^2$. For the random component coefficient vector $\underline{b}_i$, we also assume that it follows a multivariate normal distribution with zero mean and unknown variance-covariance matrix $D$. Ordinal responses $Y_{ij}$ are observed where $Y_{ij}=h$ if $\alpha'_{h-1}<\lambda'_{ij}\le\alpha'_h$. Our objectives are focused on the use of Gibbs sampler to estimate all the unknown parameters. That is we need to estimate (1) $\underline{\beta}$, (2) $\underline{b}_i$, (3) Var($\underline{b}_i$), (4) $\tau_i$ and (5) $\sigma_\tau^2$. In this model, each individual subject $i$ is assumed to have his/her own set of cut-points $\alpha'_h=\alpha_h\exp(\tau_i)$, $h=1,2,\ldots,r-1$. There are $r$ categories of responses in the model.

## 4.3 Conditional Posterior Distributions

For the Bayesian approach to the model in equation 4.3, we take a diffuse prior for $(\underline{\beta},\underline{\alpha},\lambda_{ij},\underline{b}_i,\tau_i)$. Then the joint posterior density is given by (up to proportionality)

$$\begin{aligned}\Pi(\underline{\beta},\underline{\alpha},\lambda_{ij},\underline{b}_i,\tau_i,\sigma_\tau^2\mid y_{ij}) &= c\times P(\underline{\beta},\underline{\alpha},\underline{b}_i,\tau_i)\\ &\times\prod_{i=1}^{n}\prod_{j=1}^{k}\left[\sum_{h=1}^{r}\mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1}e^{\tau_i}<\lambda_{ij}e^{\tau_i}\le\alpha_h e_i^{\tau})\right]\\ &\times\left[\frac{e^{\lambda_{ij}e_i^{\tau}-\underline{x}_{ij}^T\underline{\beta}-\underline{z}_{ij}^T\underline{b}_i}}{(1+e^{\lambda_{ij}e_i^{\tau}-\underline{x}_{ij}^T\underline{\beta}-\underline{z}_{ij}^T\underline{b}_i})^2}\right]\\ &\times|D|^{-\frac{1}{2}}\exp(-\frac{1}{2}\underline{b}_i{}^TD^{-1}\underline{b}_i)\\ &\times\frac{1}{\sigma_\tau}\exp-\frac{1}{2}(\frac{\tau_i}{\sigma_\tau})^2\times e^{\tau_i},\end{aligned}\tag{4.4}$$

where $\underline{\beta}=(\beta_0,\ \beta_1,\beta_2,\ \ldots,\beta_{p-1})^T$, $\underline{\alpha}=(\alpha_0,\ \alpha_1,\ \alpha_2,\ldots,\alpha_r\ )^T$, $\underline{b}_i=(b_0,b_1,b_2,\ldots,b_{q-1})^T$ and $q<p$. I$(\cdot)$ is the indicator function. The conditional posterior distribution of $\underline{\beta}$, denoted as

$$\left[\underline{\beta}\mid\cdot\right]=\left[\underline{\beta}\mid D,\underline{\alpha},\lambda_{ij},\underline{b}_i,\tau_i,\sigma_\tau^2,y_{ij}\right],$$

is then given by ,

$$\left[\underline{\beta} \mid \cdot\right] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) d\beta_0 d\beta_1 ... d\beta_{p-1}},$$

$$\left[\underline{\beta} \mid \cdot\right] \propto \prod_{i=1}^{n}\prod_{j=1}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \le \alpha_h e_i^{\tau})\right] \times \left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i})^2}\right], \quad (4.5)$$

The conditional density of $\lambda_{ij}$, $[\lambda_{ij} \mid \underline{\beta}, \underline{\alpha}, \underline{b}_i, D, \tau_i, \sigma_\tau^2, y_{ij}]$ , is given by,

$$[\lambda_{ij} \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) d\lambda_{ij}'},$$

$$\propto \prod_{i=1}^{n}\prod_{j=1}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \le \alpha_h e_i^{\tau})\right] \times \left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i})2}\right], \quad (4.6)$$

$$[\lambda_{ij} \mid \cdot] = \left[\lambda_{ij} \mid D, \underline{\beta}, \underline{b}_i, \tau_i, y_{ij}\right]. \quad (4.7)$$

Hence $\left[\lambda_{ij} \mid \underline{\beta}, \underline{\alpha}, \underline{b}_i, D, \tau_i, \sigma_\tau^2, y_{ij}\right] = \left[\lambda_{ij} \mid D, \underline{\beta}, \underline{b}_i, \tau_i, y_{ij}\right]$. The conditional posterior distribution of $\underline{b}_i$, $\left[\underline{b}_i \mid \underline{\beta}, \underline{\alpha}, \underline{b}_i, D, \tau_i, \sigma_\tau^2, y_{ij}\right]$, is given by

$$[\underline{b}_i \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) db_1 db_2 ... db_q}, \quad q < p$$

$$[\underline{b}_i \mid \cdot] \propto \prod_{j=1}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \le \alpha_h e^{\tau_i})\right] \times \left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i})2}\right] \times \exp(-\frac{1}{2}\underline{b}_i{}^T D^{-1}\underline{b}_i), \quad (4.8)$$

$$[\underline{b}_i \mid \cdot] = \left[\underline{b}_i \mid \underline{\beta}, \underline{\alpha}, \lambda_{ij}, \tau_i, D, y_{ij}\right]. \quad (4.9)$$

The conditional posterior density of $\underline{b}_i$ does not depends on $\sigma_\tau^2$. The conditional posterior density of Var$(\underline{b}_i) = D$, i.e. $[D \mid \underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2, y_{ij}]$, is given by

$$[D \mid \cdot] = \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \cdots \int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) d\sigma_{11} d\sigma_{12} \ldots d\sigma_{qq}}, \tag{4.10}$$

where $\sigma_{ij}, i, j = 1, 2, \ldots, q$ are the elements of $D$. Again we have

$$[D \mid \cdot] \propto |D|^{-\frac{1}{2}} \exp(-\frac{1}{2} \underline{b}_i{}^T D^{-1} \underline{b}_i). \tag{4.11}$$

Therefore the distribution of $D$ only depends on $\underline{b}_i$, i.e. $[D \mid \cdot] = [D \mid \underline{b}_i]$

For the scaling parameter $\tau_i$, the conditional posterior density is given by,

$$\begin{aligned} [\tau_i \mid \cdot] &= \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) d\tau_i} \\ [\tau_i \mid \cdot] &\propto \prod_{j=1}^{k} \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h) \mathrm{I}(\alpha_{h-1} e^{\tau_i} < \lambda_{ij} e_i^\tau \leq \alpha_h e_i^\tau) \right] \\ &\times \left[ \frac{e^{\lambda_{ij} e_i^\tau - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i}}{(1 + e^{\lambda_{ij} e_i^\tau - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i})^2} \right] \times e_i^\tau \times \exp -\frac{1}{2} (\frac{\tau_i}{\sigma_\tau})^2, \end{aligned} \tag{4.12}$$

$$[\tau_i \mid \cdot] = \left[\tau_i \mid D, \underline{\beta}, \alpha, \lambda'_{ij}, \underline{b}_i, \sigma_\tau^2, y_{ij}\right]. \tag{4.13}$$

Lastly the conditional posterior density of $\sigma_\tau^2$ is given by,

$$\begin{aligned} \left[\sigma_\tau^2 \mid \cdot\right] &= \frac{\Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij})}{\int \Pi(\underline{\beta}, \underline{\alpha}, \lambda_{ij}, \underline{b}_i, \tau_i, \sigma_\tau^2 \mid y_{ij}) d\sigma_\tau^2} \\ \left[\sigma_\tau^2 \mid \cdot\right] &\propto \frac{1}{\sigma_\tau} \exp -\frac{1}{2} (\frac{\tau_i}{\sigma_\tau})^2, \end{aligned} \tag{4.14}$$

That is, $[\sigma_\tau^2 \mid \cdot] = [\sigma_\tau^2 \mid \tau_i]$.

For the conditional posterior density of $\alpha_h$ given $\{\lambda_{ij}, \underline{\beta}, \underline{b}_i, D, \tau_i, \sigma_\tau^2, y_{ij}\}$ and $\lambda_v, \nu \neq h$, is given by (up to proportionality ),

$$\begin{aligned}[\alpha_h \mid \cdot] \quad & \propto \prod_{i=1}^{n}\prod_{j=1}^{k} \left[\mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \leq \alpha_h e_i^{\tau})\right]. \\ & + \left[\mathrm{I}(y_{ij} = h+1)\mathrm{I}(\alpha_h e^{\tau_i} < \lambda_{ij}e^{\tau_i} \leq \alpha_{h+1} e_i^{\tau})\right], \end{aligned} \tag{4.15}$$

This conditional distribution of the cut-points for each subject is also uniform. That is,

$$\begin{aligned}\alpha_h \sim \quad & \mathrm{Unif}\left[\max\{\max\{\lambda_{ij} : y_{ij} = h\}, \alpha_{h-1}\},\right. \\ & \left.\min\{\min\{\lambda_{ij} : y_{ij} = h+1\}, \alpha_{h+1}\}\right]. \end{aligned} \tag{4.16}$$

### 4.3.1 Log-concavity Conditions

In order to use adaptive rejection sampling method (ARS) to carry-out Gibbs sampling, we first need to check carefully whether the conditional posterior densities satisfy the log-concavity conditions. That is, for any density, h (say), we check whether $\frac{\partial^2 \ln h}{\partial {\beta_v}^2} < 0$ where $\beta_v$ is the $v$th parameter of interest. The conditional posterior distributions in the random effects cumulative logit model with random scaling term satisfy the log-concavity conditions, except the posterior distribution of the random scaling term ($\tau_i$). We can prove that for each $v$

$$(\mathrm{i}) \frac{\partial \ln^2 h(\underline{\beta})}{\partial {\beta_v}^2} < 0. \tag{4.17}$$

$$(\mathrm{ii}) \frac{\partial \ln^2 h(\underline{b}_i)}{\partial {b_v}^2} < 0. \tag{4.18}$$

$$(\mathrm{iii}) \frac{\partial \ln^2 h(\lambda'_{ij})}{\partial {\lambda_v}^2} < 0. \tag{4.19}$$

The proofs are almost identical to the one given in Chapter 3. It is unnecessary to repeat the proofs in this Chapter.

To check the log-concavity conditions for the scaling effect parameter $\tau_i$, we consider equation 4.12 and let

$$\begin{aligned}
[\tau_i \mid \cdot] &= h(\tau_i) \\
&= k_4 \times \prod_{i=1}^{n}\prod_{j=1}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \leq \alpha_h e_i^{\tau})\right] \\
&\times \left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i})^2}\right] \\
&\times e^{\tau_i} \times \exp -\frac{1}{2}(\frac{\tau_i}{\sigma_\tau})^2, \qquad (4.20)
\end{aligned}$$

Again for fixed $i$,

$$\begin{aligned}
\ln h(\tau_i) &= \ln k_4 + \sum_{j=1}^{k} \ln\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1}e^{\tau_i} < \lambda_{ij}e^{\tau_i} \leq \alpha_h e_i^{\tau})\right] \\
&+ \sum_{j=1}^{k}(\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i) - 2\sum_{j=1}^{k}\ln(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}) \\
&+ \left[\tau_i - \frac{1}{2}(\frac{\tau_i}{\sigma_\tau})^2\right]. \qquad (4.21)
\end{aligned}$$

$$\frac{\partial \ln h(\tau_i)}{\partial \tau_i} = \sum_{j=1}^{k}\lambda_{ij}e^{\tau_i} - 2\sum_{j=1}^{k}\lambda_{ij}e^{\tau_i}\left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}\right] + 1 - \frac{\tau_i}{\sigma_\tau^2}, \qquad (4.22)$$

and

$$\begin{aligned}
\frac{\partial \ln^2 h(\tau_i)}{\partial \tau_i^2} &= \sum_{j=1}^{k}\lambda_{ij}e^{\tau_i} - 2\sum_{j=1}^{k}\lambda_{ij}e^{\tau_i}\left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}\right] \\
&- 2\sum_{j=1}^{k}(\lambda_{ij}e^{\tau_i})^2\left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}\right] \\
&- 2\sum_{j=1}^{k}(\lambda_{ij}e^{\tau_i})^2\left[\frac{e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}}{\left(1 + e^{\lambda_{ij}e^{\tau_i} - \underline{x}_{ij}^T\underline{\beta} - \underline{z}_{ij}^T\underline{b}_i}\right)^2}\right] - \frac{1}{\sigma_\tau^2}. \qquad (4.23)
\end{aligned}$$

The log-concavity condition does not hold in this case. Therefore to generate a sample variate of $\tau_i$ we should use the Metropolis sampling method or alternatively we could well use the 'Adaptive Rejection Metropolis Sampling' method (ARMS) by Gilks *et al.* (1995). The ARMS method deals with a distribution $g$ when $g$ is nearly log-concave. Lastly, the Odell-Feivison (1966) technique is used to generate elements of random matrix $D$ and $\sigma_\tau^2$ will be generated by using an inverse gamma distribution.

It is not difficult to prove log-concavity conditions of h($\underline{\beta}$) and h($\underline{b}_i$) for any design matrix. The proofs are identical to the one given in the previous two chapters.

## 4.4 Simulation Studies

To carry out a simulation study for our proposed model, we consider the following cumulative logit model with random effects:

$$\operatorname{logit} \Pr(Y_{ij} \leq h \mid \underline{b}_i) = \alpha_h \exp(\tau_i) - (\beta_0 + \beta_1 t + \beta_2 x_i + \beta_3 t x_i + b_{0i} + b_{1i} t) \quad (4.24)$$

where $x_i = 0$ for half of the population and 1 for the remainder and t = -3, -2, -1, 0, 1, 2, 3. The fixed effects coefficients were set at $\beta_0 = +3, \beta_1 = -2.5, \beta_2 = +2.5$ and $\beta_3 = -1.0$. The random scaling terms $\tau_i$ is simulated according to the normal distribution with mean 0 and unit variance. The random effects distribution is simulated with

$$\operatorname{Var}(\underline{b}_i) = D = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.001 \end{bmatrix} \quad (4.25)$$

Ordinal response data are simulated according to equation 4.24 with three cut-points. Each data set was comprised of n = 100 (subjects) clusters of size j = 7 for each $i$th subject.

To perform Gibbs sampling, we run the chain for each parameter of interest for 11,000 times and discard the initial 1000 values as burn-in. Table 4.1 shows the overall results of one typical run using one simulated data set. The results in Table 4.1 indicate that the Gibbs sampler scheme gives satisfactory results.

Table 4.1. Results of Simulation Studies

| Parameter | True value | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| $\beta_0$ | 3.00 | 2.9671 | 2.9523 | 0.2801 | 2.1563 | 5.0371 |
| $\beta_1$ | -2.50 | -2.5937 | -2.6735 | 0.1978 | -3.1378 | -2.0674 |
| $\beta_2$ | 2.50 | 2.4965 | 2.5271 | 0.7239 | 1.0935 | 6.9673 |
| $\beta_3$ | -1.00 | -1.0916 | -1.0734 | 0.4109 | -2.9067 | 0.1532 |
| Var($b_0$) | 1.00 | 0.9323 | 0.5925 | 0.5815 | 0.1734 | 5.9362 |
| Var($\tau_i$) | 1.00 | 0.8975 | 0.6946 | 0.6012 | 0.0938 | 6.9015 |

In Table 4.1, estimates using the ARS method are generally close to the true values. The potential scale reduction factor (PSRF) R=1.0097 which indicates that the Gibbs sampler scheme converged very well. Since this is a simulation study and the value of R is close to 1. Therefore we assumed with confidence that each individual series is well converged.

## 4.5 Analysis of BT Ordinal Data

As we have mentioned earlier in section 3.6 concerning the BT experimental data, the data that we analyze in this chapter are typically of a five point scale graded from 'Bad' to 'Excellent'. This belongs to the ordinal response scale. In the following we present the final results of our analysis for E198 experiment using random effects cumulative logit model with subject-specific scaling term. As in the simulation study, we generated 10,000 random variates for each parameter after 1000 burn-in values. We need to estimate the intercept term, parameters for the 32 subject effects, parameters for the 7 condition effects (condition 1 is aliased), random effects variance and variance for the random scaling effects. There are altogether 42 parameters to be estimated. The cut-points are of less interest since we are focusing on the estimation of the regression parameters $\underline{\beta}$ for the random scaling model and the variance of the random scaling effects. In Table 4.2 we show the results of our estimates.

Table 4.2. Results of the BT (E198) experiment. The estimates are reported along with the mean, the standard error, the standard error of batching mean, the lower 2.5*th*, ($P_{0.025}$) and the upper 97.5*th*, ($P_{0.975}$) percentiles.

| para. | mean | s. d. | s. d.[1] | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|---|
| Intercept | 7.8262 | 0.6343 | 0.0452 | 7.0977 | 8.5378 |
| Subject 1 | -0.5872 | 0.7903 | 0.0562 | -1.8921 | 0.7823 |
| Subject 2 | -0.4781 | 0.7599 | 0.0265 | -1.6418 | 0.6448 |
| Subject 3 | -0.5198 | 0.8904 | 0.0301 | -2.0999 | 1.0593 |
| Subject 4 | 0.4961 | 0.8478 | 0.0451 | -0.9315 | 1.9253 |
| Subject 5 | -1.0840 | 0.8794 | 0.0339 | -2.6304 | 0.4088 |
| Subject 6 | 0.8884 | 0.7821 | 0.0278 | -0.3232 | 2.1049 |
| Subject 7 | -0.9763 | 0.8229 | 0.0421 | -2.2704 | 0.3104 |
| Subject 8 | -1.2185 | 0.8308 | 0.0329 | -2.6200 | 0.1212 |
| Subject 9 | 0.2185 | 0.9226 | 0.0420 | -1.4239 | 1.9904 |
| Subject 10 | 0.8725 | 0.8077 | 0.0297 | -0.3988 | 2.1941 |
| Subject 11 | 1.2014 | 0.8303 | 0.0411 | -0.0768 | 2.5283 |
| Subject 12 | -1.4146 | 0.8188 | 0.0388 | -2.7351 | -0.1184 |
| Subject 13 | -1.3004 | 0.8132 | 0.0225 | -2.6423 | -0.0268 |
| Subject 14 | 1.5976 | 0.8887 | 0.0261 | 0.0131 | 3.1558 |
| Subject 15 | 0.1183 | 0.8348 | 0.0391 | -1.2359 | 1.5233 |
| Subject 16 | -0.6576 | 0.7705 | 0.0203 | -1.8659 | 0.5127 |
| Subject 17 | 2.3406 | 0.8466 | 0.0318 | 0.9110 | 3.7070 |
| Subject 18 | 0.8946 | 0.8418 | 0.0460 | -0.5468 | 2.2594 |
| Subject 19 | 0.8963 | 0.8313 | 0.0433 | -0.4327 | 2.3083 |
| Subject 20 | -1.3037 | 0.7896 | 0.0308 | -2.5749 | -0.0897 |
| Subject 21 | -2.5688 | 0.9356 | 0.0486 | -4.1598 | -0.8066 |
| Subject 22 | 0.5520 | 0.8115 | 0.0453 | -0.7673 | 1.8607 |
| Subject 23 | -2.8102 | 0.7768 | 0.0329 | -3.9664 | -1.6266 |
| Subject 24 | 0.0877 | 0.8394 | 0.0214 | -1.2664 | 1.4608 |

Table 4.2. (Continued) Results of the BT (E198) experiment. The estimates are reported along with the mean, the standard error, the standard error of batching mean, the lower 2.5*th*, ($P_{0.025}$) and the upper 97.5*th*, ($P_{0.975}$) percentiles.

| para. | mean | s. d. | s. d.[1] | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|---|
| Subject 25 | 3.7748 | 0.8877 | 0.0327 | 2.2505 | 5.3246 |
| Subject 26 | -0.0814 | 0.8158 | 0.0221 | -1.3977 | 1.2331 |
| Subject 27 | 3.0671 | 0.8658 | 0.0331 | 1.5785 | 4.5042 |
| Subject 28 | -0.0026 | 0.8329 | 0.0402 | -1.3369 | 1.3784 |
| Subject 29 | -1.4563 | 0.8232 | 0.0403 | -2.7404 | -0.1160 |
| Subject 30 | -0.7020 | 0.8060 | 0.0392 | -1.9663 | 0.5572 |
| Subject 31 | -0.2009 | 0.8167 | 0.0365 | -1.4259 | 1.2204 |
| Subject 32 | -7.4838 | 0.7826 | 0.0289 | -8.7121 | -6.2673 |
| Cond 1 | 0 (aliased) | | | | |
| Cond 2 | -0.3327 | 0.7181 | 0.0301 | -1.3103 | 0.6919 |
| Cond 3 | -0.1603 | 0.7386 | 0.0288 | -1.2079 | 0.8585 |
| Cond 4 | -6.8902 | 0.7231 | 0.0231 | -7.8772 | -5.9009 |
| Cond 5 | -0.0968 | 0.7770 | 0.0403 | -1.2380 | 1.0788 |
| Cond 6 | -2.8247 | 0.7204 | 0.0392 | -3.7741 | -1.8361 |
| Cond 7 | -6.3403 | 0.7158 | 0.0457 | -7.3050 | -5.3472 |
| Cond 8 | -7.3907 | 0.7211 | 0.0313 | -8.3408 | -6.4186 |
| Var($b_0$) | 0.9418 | 0.4581 | 0.0521 | 0.6245 | 1.5617 |
| Var($\tau_i$) | 1.8987 | 0.5572 | 0.0539 | 1.1425 | 2.3527 |

1: This is standard deviation of the batching means

As we can see from Table 4.2, the batching standard deviations are very small relative to overall sample standard deviations in each chain. The number of sample points in each batch is 500. There are 20 batches. The PSRF R=1.0056 for Table 4.2. This indicates that that convergence is good in each of the Gibbs sampling scheme. The last two columns in each of Table 4.2 show the 2.5% and 97.5% quantile values. In the following, we present the empirical posterior distribution of the intercept, condition 8 effect, random effect variance and variance of the random scaling terms ($\tau_i$).

Figure 4.11: Empirical Posterior Distribution of Intercept

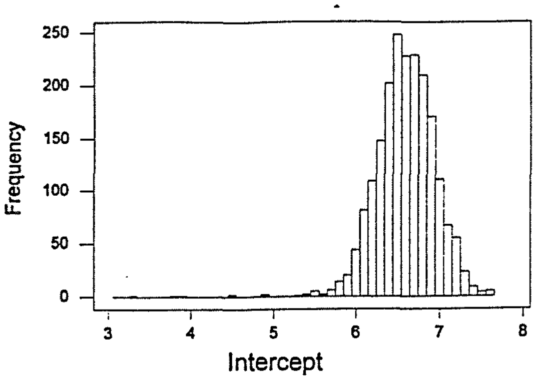
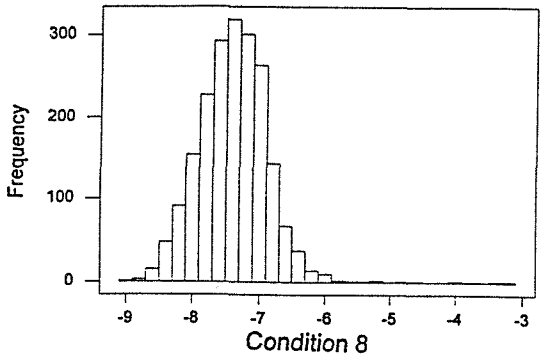
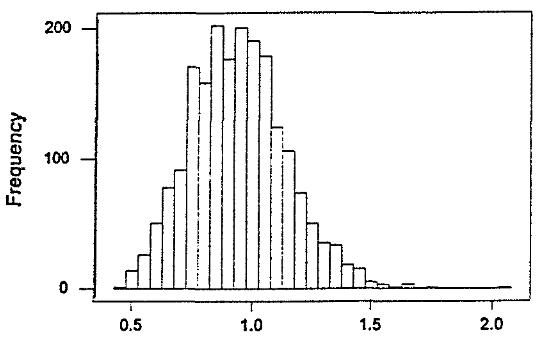



Figure 4.12: Empirical Posterior Distribution of Condition 8



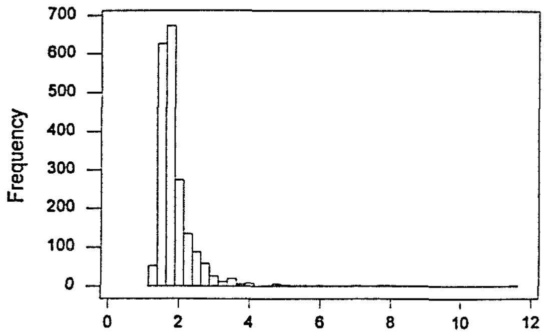

83

Figure 4.13: Empirical Posterior Distribution of R. E. Variance




Random Effect Variance

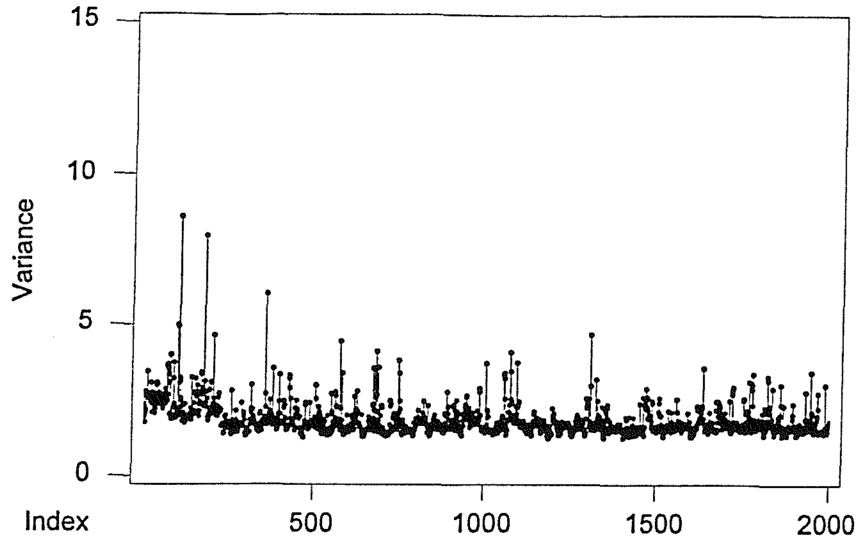
Figure 4.14: Empirical Posterior Distribution of Variance of Random Scaling Term




Random Scaling Variance

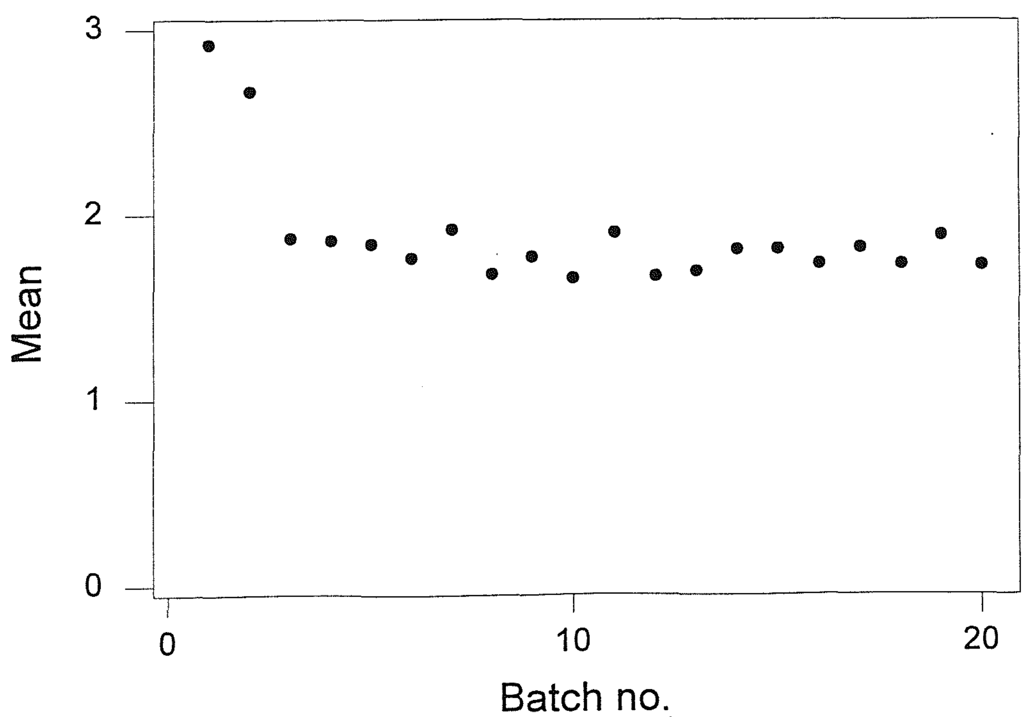
The random scaling term $\tau_i$ is generated by the Metropolis scheme which differs from the rest of the other parameters. All the other parameters are used the adaptive rejection sampling (ARS) scheme. The ARS scheme is proved to be converged in previous chapters. It will be of interest to show a time series plot of the variance of the generated random scaling terms using the Metropolis scheme. The series is taken from the last 2,000 values of the original series of size $n = 10,000$.

Figure 4.15: Time Series Plot of the Generated Random Scaling Variance




To show that the series is indeed convergent, we further show the batch mean plot of the series in Figure 4.16. The batch means are calculated from the last 2,000 generated values. Each batch is of size 100 and there are 20 batches. As we can see from Figure 4.16 that batch mean indeed fluctuats around the mean level of 1.90.

Figure 4.16: Time Series Plot of Batch Means of Random Scaling Variances



### 4.5.1 Posterior-Predictive Probability Distributions

In this section we show further the predictive probabilities for future events from the posterior-predictive density using the random scaling model. Again we compute the posterior-predictive distribution of scoring a bad telephone connection of subject 8 and subject 16 at each condition level. The following Tables show the posterior-predictive distribution of subject 8 and subject 16. These figures are calculated based on the simulated values of $\underline{\beta}^{(j)}, j = 1, 2, \ldots, m$. The sample size $m$ is also equal to 2,000 and the sample is taken from the last 2,000 values of the entire simulated Markov chain.

Table 4.3 Predictive probabilities of having a bad telephone connection for subject 8 at each given conditions (E198 experiment, random scaling model).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0014 | 0.0419 | 0.0002 | 0.0043 |
| Cond 2 | 0.0019 | 0.0393 | 0.0003 | 0.0063 |
| Cond 3 | 0.0017 | 0.0376 | 0.0003 | 0.0054 |
| Cond 4 | 0.5087 | 0.4112 | 0.1926 | 0.8277 |
| Cond 5 | 0.0016 | 0.0371 | 0.0002 | 0.0053 |
| Cond 6 | 0.0228 | 0.1359 | 0.0040 | 0.0727 |
| Cond 7 | 0.3890 | 0.4009 | 0.1205 | 0.7101 |
| Cond 8 | 0.6174 | 0.4023 | 0.2853 | 0.8845 |

Table 4.4 Predictive probabilities of having a bad telephone connection for subject 16 at each given conditions (E198 experiment, random scaling model).

| Conditions | $\Pr(y^* = 1 \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0005 | 0.0277 | 0.0001 | 0.0017 |
| Cond 2 | 0.0007 | 0.0249 | 0.0001 | 0.0023 |
| Cond 3 | 0.0006 | 0.0221 | 0.0001 | 0.0019 |
| Cond 4 | 0.2815 | 0.3779 | 0.0701 | 0.6096 |
| Cond 5 | 0.0006 | 0.0236 | 0.0001 | 0.0020 |
| Cond 6 | 0.0079 | 0.0811 | 0.0013 | 0.0257 |
| Cond 7 | 0.1925 | 0.3381 | 0.0395 | 0.4697 |
| Cond 8 | 0.3805 | 0.4028 | 0.1068 | 0.7063 |

The figures in Table 4.3 and 4.4 are different from those calculated using the random effects cumulative logit model. Again one of the advantage of Bayesian statistical modelling is one can easily obtain the predictive probabilities of certain events.

### 4.5.2 Latent Residuals Analysis

Finally for model checking, we use again the method of latent residuals introduced by Albert and Chib (1995). Latent residuals are generated in the same way as in Chapter 3. Latent residuals are generated for each subject $i$. That is if 4.4 is the correct model for the data, then for each response category h,

$$\lambda'_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{b}_i + \varepsilon_{ij} \tag{4.26}$$

where $\lambda'_{ij} = \lambda_{ij} e^{\tau_i}$ are the latent variables for the h category and $\alpha'_{h-1} < \lambda'_{ij} < \alpha'_h$. $\alpha'_{h-1}$ and $\alpha'_h$ are the corresponding cut-points for each subject $i$. Latent variable $\lambda'_{ij}$ follows a logistic distribution with mean $\underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{b}_i$ and unit variance. $\lambda'_{ij}$ can be generated directly from the ARS method for given values of $\underline{\beta}^{(t)}$ and $\underline{b}_i{}^{(t)}$ at the $tth$ stage of iteration of our Gibbs sampler scheme as we have done in Section 5.

Therefore latent residuals $\varepsilon_{ij}$ are defined as

$$\varepsilon_{ij} = \lambda'_{ij} - \underline{x}_{ij}^T \underline{\beta} - \underline{z}_{ij}^T \underline{b}_i \tag{4.27}$$

Latent residuals $\varepsilon_{ij}$ are a priori a random sample from a standard logistic distribution.

To show that the generated latent residuals of our model followed a standard logistic distribution, a Kolmogorov-Smirnov test is being used again to test whether $\varepsilon_{ij}$ do follow a standard logistic distribution. To conduct the test a random sample of size 30 is being picked from the 5000 generated latent residuals each time and we obtain the necessary $D_n$ test statistic as well as Prob($D \leq D_n$). The test is repeated for 50, 100 and 1000 times.

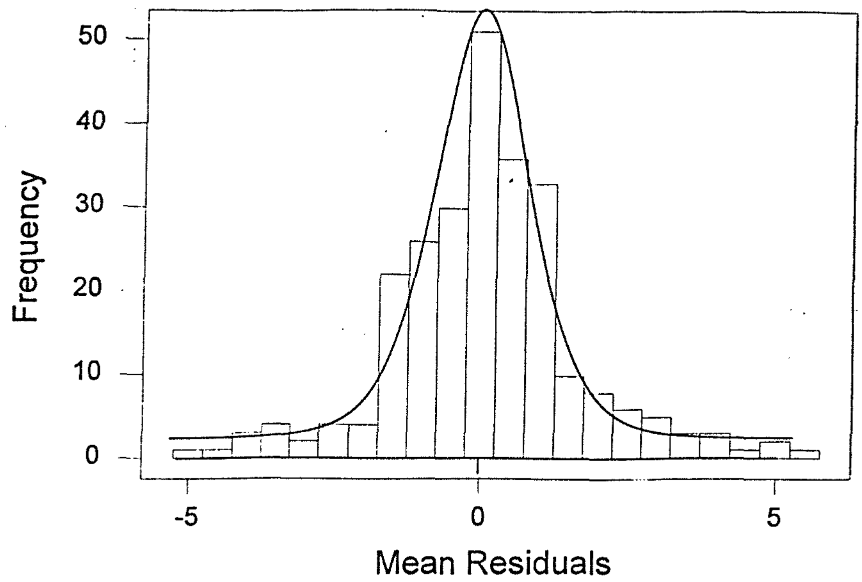
Table 4.5. Results of K-S Test of latent residuals

(E198 Experiment, random scaling effects model)

| Sample size (n) | No. of runs (m) | Mean $D_n$ | Mean Pr($D \leq D_n$) |
|---|---|---|---|
| 30 | 50 | 0.1508 | 0.5789 |
| 30 | 100 | 0.1496 | 0.5134 |
| 30 | 1000 | 0.1531 | 0.5543 |

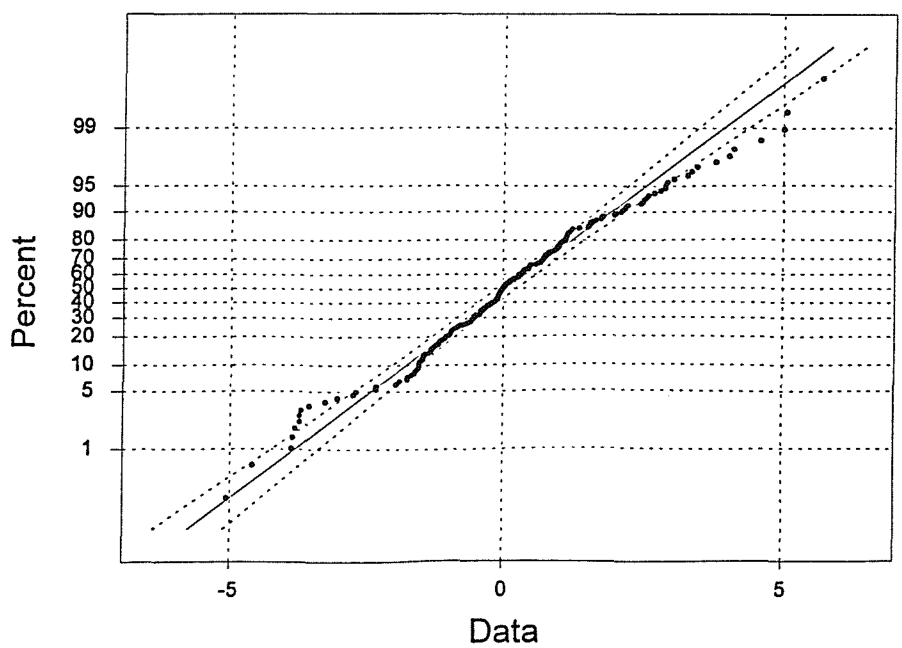
The histogram of mean residuals for each observation for each subject are shown in Figure 4.7. The empirical distribution of the 256 mean latent residuals is superimposed by a standard logistic distribution curve. From looking at Figure 4.7, the empirical distribution is in close agreement with the underlying theoretical distribution.

Figure 4.17: Histogram of Mean Residuals of Random Scaling Model



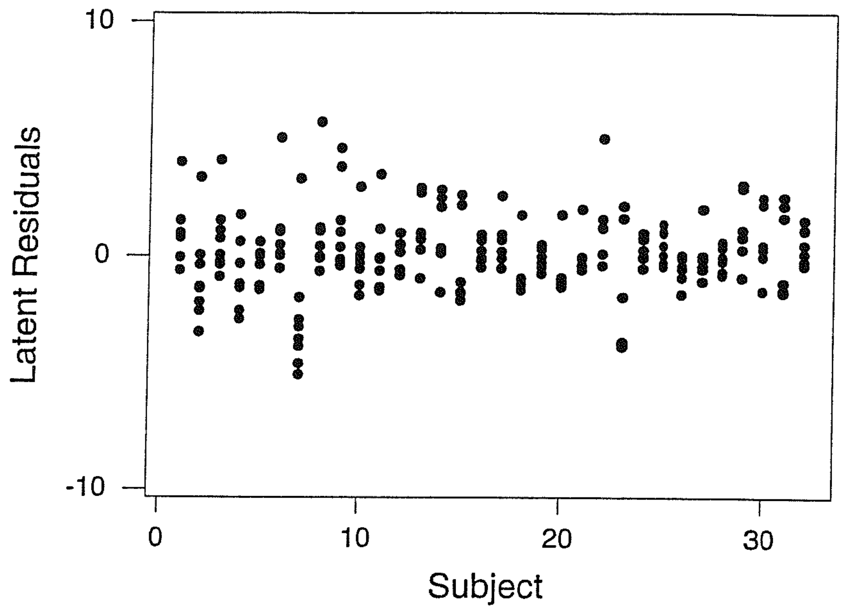

We further show the logistic probability plot of the mean latent residuals in Figure 4.18. The logistic probability plot shows that in that the residual scores are scattered around a straight line in the middle 90% of the data. This strongly indicates that the residuals follow a standard logistic distribution which agrees with our basic assumption.

Figure 4.18: Logistic Probability Plot of Mean Residuals of Random Scaling Model



A plot of mean latent residuals plot against subject is shown in Figure 4.19 below. This plot also indicates that the E198 data are better fitted by the random scaling model. Majority of the mean latent residuals are within the ±3.0 limits. No particular trends or patterns are found in this plot.

Figure 4.19: Mean Latent Residual Plot against Subject of Random Scaling Model




## 4.6 Conclusions

The random effects cumulative logit model with subject-specific scaling effects is an extension of the cumulative logit model with random effects. This model caters for the experiments involved with repeated measurement where the data observed are ordinal categorical data. It is assumed that each subject under investigation may have their own scale of cut-points. Our empirical results have shown that this method of modelling ordinal data of this nature is possible. The BUGS software package has not yet developed programs to cater for problems of this kind. It is possible that most practitioners do not realize that the subjects under investigation has its own scale of cut-points. One advantage of using ARS is that latent data and associated residuals can be computed within an ARS-based sampler for the ordinal logistic random effects model.

Estimation of the parameters in the cumulative logit model with random effects can be done by using generalized estimating equations (GEE). Zeger *et al.* (1988) demonstrate how this is done in fitting the logistic regression model for binary responses. Using GEE to fit the model involves the use of a working correlation matrix. Details of the GEE approach may be found in Zeger *et al.* (1988).

# Chapter 5

# Continuation-ratio Logit Model

## 5.1 Introduction

Models for the cumulative probability of category $h$, $\gamma_{ij,h} = \Pr(Y_{ij} \leq h)$, have already been introduced in Chapter 3. The cumulative logit model is defined in equation (3.1). In this thesis models based on the cumulative probability are referred to as cumulative link models. When we discuss a particular member of the class of cumulative link models, the word link is replaced by the name of the particular link function being used. However, McCullagh (1980) proposed the name " proportional odds" for the cumulative logit model and "proportional hazards" for the cumulative complementary log-log model.

All models based on cumulative probability and cumulative logit imply stochastic ordering. The cumulative logit and cumulative probits are invariant due to a combination of the symmetric nature of the link functions and the simple form of the cumulative probability. All the cumulative link models are invariant to contiguous category collapsing.

Another probability that has been discussed in the literature (Agresti, 1996) is the continuation-ratio probability.

$$\begin{aligned}\gamma_{ij,h} &= \Pr(Y_{ij} = h \mid Y_{ij} \geq h) \\ &= \frac{\pi_{ij,h}}{\pi_{ij,h} + \cdots + \pi_{ij,r}}, \qquad (h = 1, 2, ...r)\end{aligned} \tag{5.1}$$

Any appropriate link function may be used to form a model with the continuation-ratio probability. The *continuation − ratio logits* contrast each category with a grouping of categories from higher levels of the response scale; that is,

$$\log\left[\frac{\pi_1}{\pi_2 + \cdots + \pi_r}\right], \log\left[\frac{\pi_2}{\pi_3 + \cdots + \pi_r}\right], \ldots, \log\left[\frac{\pi_{r-1}}{\pi_r}\right] \tag{5.2}$$

For example, what will be referred to as the continuation-ratio logit model of the second type for ordinal response with regression parameter vector $\underline{\beta}$ is of the following form:

$$\log\left[\frac{\pi_{ij,h}}{\pi_{ij,h+1} + \cdots + \pi_{ij,r}}\right] = \alpha_h - \underline{x}'_{ij}\underline{\beta} \tag{5.3}$$

or equivalently

$$\log\left[\frac{\pi_{ij,h}}{1 - \gamma_{ij,h}}\right] = \alpha_h - \underline{x}'_{ij}\underline{\beta} \tag{5.4}$$

Similarly there are the continuation-ratio complementary log-log and continuation-ratio probit models when the link function is the complementary log-log and probit respectively. The continuation-ratio link models do not have the same appeal to an underlying continuum as in the case with the cumulative link models. All continuation-ratio link models define strict stochastic ordering but they are not, in general, invariant to the collapsing of the contiguous categories.

## 5.2 Relationship between Continuation-ratio Model and Model for Binary Data

To estimate the parameters of a continuation-ratio logit model using the Gibbs sampler, we can use the estimation method of chapter 2 again here. It is easy to show that the likelihood function for continuation-ratio link models can be split into $r-1$ independent binomial likelihood functions. To demonstrate, consider the case of an ordinal response having four categories ($r$ =4). The likelihood $L$ for the $ij$th observation is proportional to

$$L \quad \propto \quad \pi_{ij,1}^{y_{ij,1}} \times \pi_{ij,2}^{y_{ij,2}} \times \pi_{ij,3}^{y_{ij,3}} \times \pi_{ij,4}^{y_{ij,4}}, \tag{5.5}$$

where $y_{ij,h} = 1$ if the ordinal response $y_{ij} = h$ and $y_{ij} = 0$ otherwise. This may be written as

$$\begin{aligned} L \quad \propto \quad & \pi_{ij,1}^{y_{ij,1}} (\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4})^{y_{ij,2}+y_{ij,3}+y_{ij,4}} \\ \times \quad & \left(\frac{\pi_{ij,2}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right)^{y_{ij,2}} \times \left(\frac{\pi_{ij,3} + \pi_{ij,4}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right)^{y_{ij,3}+y_{ij,4}} \\ \times \quad & \left(\frac{\pi_{ij,3}}{\pi_{ij,3} + \pi_{ij,4}}\right)^{y_{ij,3}} \times \left(\frac{\pi_{ij,4}}{\pi_{ij,3} + \pi_{ij,4}}\right)^{y_{ij,4}} \end{aligned} \tag{5.6}$$

Since $\sum_{h=1}^{4} \pi_{ij,h} = 1$, this is also the joint likelihood for one observation from each of the three independent binomial distributions,

(i) $\text{Bin}(1, \pi_{ij,1})$

(ii) $\text{Bin}\left(y_{ij,2} + y_{ij,3} + y_{ij,4}, \frac{\pi_{ij,2}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right)$

(iii) $\text{Bin}\left(y_{ij,3} + y_{ij,4}, \frac{\pi_{ij,3}}{\pi_{ij,3} + \pi_{ij,4}}\right)$

A continuation-ratio link model is in this case

$$\begin{aligned} \text{Link}(\pi_{ij,1}) &= \alpha_1 - \underline{x}'_{ij}\underline{\beta} \\ \text{Link}\left(\frac{\pi_{ij,2}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right) &= \alpha_2 - \underline{x}'_{ij}\underline{\beta} \\ \text{Link}\left(\frac{\pi_{ij,3}}{\pi_{ij,3} + \pi_{ij,4}}\right) &= \alpha_3 - \underline{x}'_{ij}\underline{\beta} \end{aligned}$$

So if the continuation-ratio link model is considered to have three levels then each level models one of the binomial probabilities in the expanded likelihood. This means that we can use the method described in chapter 2 for fitting logistic regression models to fit continuation-ratio link models at each level. However, due to fact that the parameters of the continuation-ratio model are not invariant to the collapsing of the contiguous categories, therefore it is difficult to estimate the random effects components. The structure of the original random effects affecting each category has been severely distorted by the collapsing of the contiguous categories. That is, if we let $y_{ij,h} = 1$ if the ordinal response $y_{ij} = h$ and $y_{ij,h} = 0$ if $y_{ij,h} > h$. However we could easily includes random effects if required.

## 5.3 Simulation Studies of Continuation-Ratio Model

To carry out simulation work for the continuation-ratio model, it is not appropriate to simulate a certain set of ordinal data (as we did in Chapter 3 and 4) with a set of fixed parameters because we know that the parameters of the continuation-ratio model at each level will be different and it is difficult to guess the expected values of the parameters in the model at each level.

One way to demonstrate the estimation works is to consider the simulation model in Chapter 3 where the ordinal data are generated according to the following cumulative logit model with random effects:

$$\text{logit}(Y_{ij} \leq h \mid \underline{b}_i) = \lambda_h - (\beta_0 + \beta_1 t + \beta_2 x_i + \beta_3 t x_i + b_{0i} + b_{1i} t) \tag{5.7}$$

where $x_i = 0$ for half of the population and 1 for the remainder and t = -3, -2, -1, 0, 1, 2, 3. The fixed effects coefficients were set at $\beta_0 = +3, \beta_1 = -2.5, \beta_2 = +2.5$ and $\beta_3 = -1.0$. The random effects distribution is simulated with

$$\text{Var}(\underline{b}_i) = D = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.001 \end{bmatrix} \tag{5.8}$$

This model has three cut-points ($h = 1, 2, 3$) and four ordered categories. As in Chapter 3, each data set comprised of n = 100 (subjects) clusters of size j = 7 for each $i$th subject. To perform the Gibbs sampler, the second, the third and the fourth categories are merged into one category. This is a continuation-ratio on its own right. The estimated parameters will be close to the original parameters. The underlying structure of the model is a random effects cumulative logit model. Now the random effects which are used to generate the ordinal data are buried as 'noises' in the continuation-ratio model. So the final estimates will not get as close as to the original parameters. Some errors in estimation are bound to occur due to random effects superimposed in the model. The following is the final results of our estimates by the ARS method. As usual each chain in the Gibbs sampler is run for 11,000 times and we discard the initial 1000 values as burn-in. Table 5.1 shows the details of the overall results of one typical run using one simulated data set.

Table 5.1. Results of Simulation Studies of continuation ratio model

| Parameter | True value | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| $\beta_0$ | 3.0 | 3.4509 | 3.4650 | 0.4391 | 1.0407 | 5.3380 |
| $\beta_1$ | -2.50 | -2.3171 | -2.3075 | 0.2581 | -3.2602 | -0.8935 |
| $\beta_2$ | 2.50 | 2.1877 | 2.1221 | 0.9522 | 1.1531 | 5.8204 |
| $\beta_3$ | -1.00 | -0.9190 | -0.8918 | 0.5154 | -2.6930 | 1.0873 |

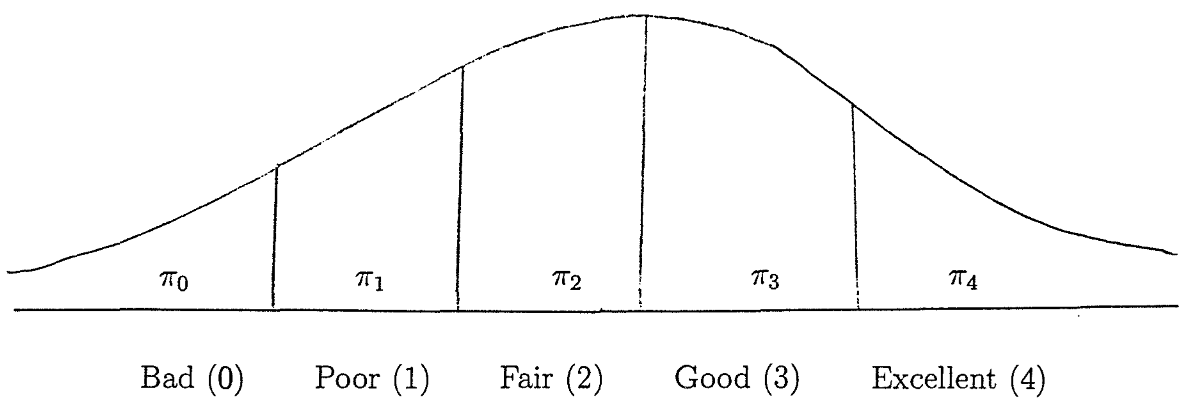
The potential scale reduction factor (PSRF) R=1.0149 in this simulation case. That means the Gibbs sampler scheme converged well. As expected, there is an average about 10% underestimates for all the parameters. This confirmed the fact stated at the beginning of this section.

We could well perform another simulation for merging other categories. However as this is only a small chapter in illustrating continuation-ratio model, the above simulation work will suffice to show parameter estimation of continuation-ratio model.

## 5.4 Continuation-Ratio Model to Fit The Telephone Connection Quality Data

As we have mentioned in Chapter 3, the response in British Telecom experiments on transmission assessment is a five-point scale, graded subjectively from "Excellent " to "Bad" (see Figure 5.1). This and similar variations are the recommended scales being used in telecommunications work (CCITT 1992).

Figure 5.1: Response Scale of Ordinal Data




In Figure 5.1, $\pi_r$, $r = 0, 1, 2, 3, 4,$ is the probability of $Y_{ij} = r$ and $\gamma_{ij,r}$ is the cumulative probability of $Y_{ij} \leq r$.

In this chapter a continuation-ratio model is used to fit the BT telephone connection quality data. The continuation-ratio model is a model which, in general, models any response of an ordered categorical nature. This implies that the model would be suitable to cope with many other ordered categorical scales.

The response is not continuous but rather discrete with only five possible values. More information is contained within the ordered structure of the categories. The categories are strictly increasing from "Bad" to " Excellent". Also relevant is the fact that these categories are not fixed but may be thought of as arbitrary coding of some underlying continuum. This underlying continuum is the unmeasurable subjective response such as "standard of transmission".

As a result of these considerations, it is the probability of a response falling into a certain category ($\pi_i$) which is the focus of the modelling procedure. Associated with this are the cumulative probabilities ($\gamma_{ij,r}$), the probabilities of the response falling in a certain category or below it.

Here we propose to use the continuation-ratio model to model the the E198 data set using the expanded likelihood with one common set of regression parameters, but different intercepts (cutpoints). In the following we present the final results of our analysis. As in simulation, 10,000 random variates are generated for each parameter after 1000 burn-in values. The results of the final estimates are shown in Table 5.2.

Table 5.2. Results for Continuation-ratio model of E198 experiment. The estimates are reported along with the mean, the standard error, the standard error of batching mean, the lower $2.5th, (P_{0.025})$ and the upper $97.5th, (P_{0.975})$ percentiles.

| para. | mean | s. e. | s. e.[1] | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|---|
| Intercept | -8.9036 | 0.6313 | 0.0074 | -9.4885 | -8.3079 |
| Cond 1 | aliased | - | - | - | - |
| Cond 2 | 0.2835 | 0.7280 | 0.0106 | -0.6140 | 1.1834 |
| Cond 3 | 0.3856 | 0.7130 | 0.0208 | -0.6078 | -1.3223 |
| Cond 4 | 7.3293 | 0.7301 | 0.0116 | 6.3213 | 8.3255 |
| Cond 5 | 0.1846 | 0.7302 | 0.0417 | -0.8195 | 1.1492 |
| Cond 6 | 2.6563 | 0.7195 | 0.0301 | 1.6598 | 3.5756 |
| Cond 7 | 6.4176 | 0.7195 | 0.0113 | 5.4333 | 7.3969 |
| Cond 8 | 8.0449 | 0.7611 | 0.0205 | 7.0102 | 8.9981 |
| Cut 1 | 0 | - | - | - | - |
| Cut 2 | 3.3255 | 0.6902 | 0.0753 | 2.4298 | 5.9761 |
| Cut 3 | 6.7925 | 0.6384 | 0.0114 | 5.4651 | 7.8235 |
| Cut 4 | 9.3560 | 0.7098 | 0.0374 | 7.5261 | 10.7382 |

1: This is the standard error of the batching means.

As we can see from Table 5.2, the standard errors of batching means are very small relatively to overall standard errors of the sample means in each chain. The number of sample points in each batch is 500. There are 20 batches. The value of PSRF for Table 5.2 is equal to 1.0217. This indicates that convergence is good in each of the Gibbs sampling schemes. The last two columns of Table 5.2 give the usual 2.5% and 97.5% quantile values as shown in the last three chapters.

We can in fact compare our results with those obtained by Lewis *et al.* (1992) who used the GLIM software package to fit the continuation-ratio model to the E198 data. However, they only included 16 subjects in the model whereas we used 32 subjects and they included further covariates in addtion to those that we have used here. Table 5.3 is extracted from their results. It is noted that the estimates for the cut-points and transmission conditions are similar to our results.

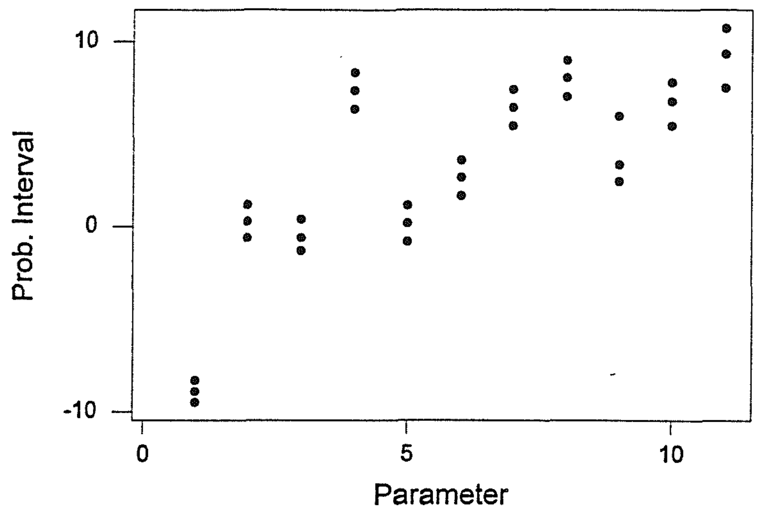
Table 5.3. Results Continuation-ratio model of E198 experiment by Lewis *et al.* (1992)[2]

| parameter | estimate | s. e. |
|---|---|---|
| Intercept | -12.68 | 1.488 |
| Cond 1 | aliased | - |
| Cond 2 | 0.5848 | 0.7317 |
| Cond 3 | 1.067 | 0.7392 |
| Cond 4 | 8.924 | 1.043 |
| Cond 5 | 0.9571 | 0.7349 |
| Cond 6 | 4.427 | 0.7999 |
| Cond 7 | 7.506 | 0.9487 |
| Cond 8 | 10.24 | 1.125 |
| Cut 1 | 0 | - |
| Cut 2 | 3.764 | 0.5271 |
| Cut 3 | 6.992 | 0.7302 |
| Cut 4 | 10.27 | 0.8831 |

2:This table is extracted from Table C.1 of Lewis *et al.* (1992).

The probability intervals and the estimated mean values of the parameter in Table 5.2 are presented in Figure 5.2

Figure 5.2 : Plots of Probability Intervals of Parameters




The empirical posterior distributions of the intercept term, effects of condition 2, condition 4 and condition 6 are shown in the following graphs. The sample size in each histogram is equal to 2,000. These values are taken from the last portion of the original simulated chain (n=10,000). Hopefully a sample of 2,000 is good enough to show the true posterior distribution.

Figure 5.3: Empirical Posterior Distribution of Intercept

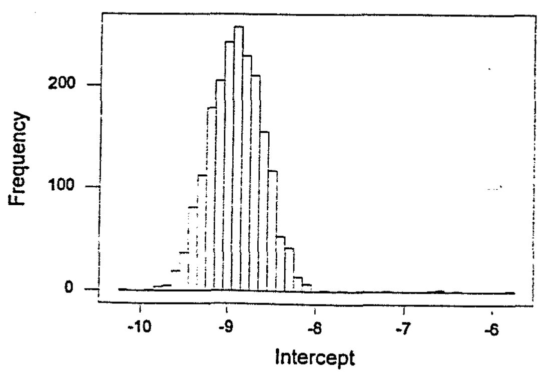



Figure 5.4: Empirical Posterior Distribution of Condition 2

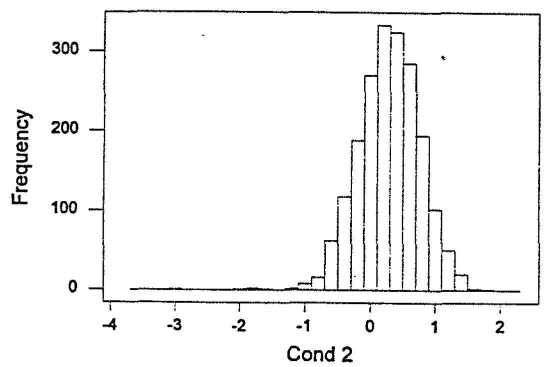


Figure 5.5: Empirical Posterior Distribution of Condition 4

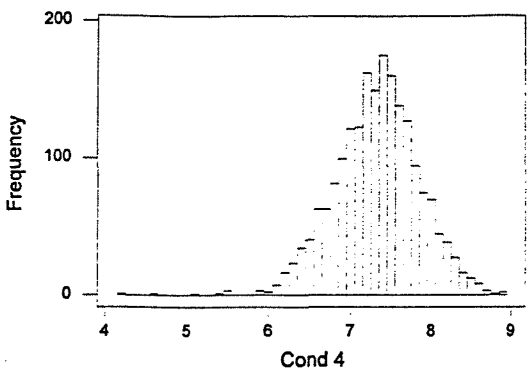
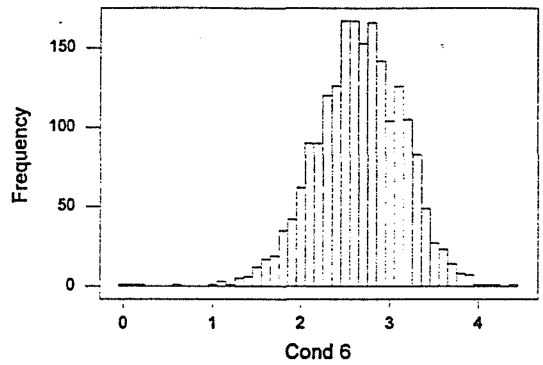



Figure 5.6: Empirical Posterior Distribution of Condition 6




### 5.4.1 Posterior-Predictive Distributions

For the continuation-ratio model of the E198 experiment, we compute the posterior-predictive distribution of the first level for subject 10 and subject 12 at each condition level. There are eight conditions. This is the posterior-predictive distribution of having a bad connection. Here 'bad' is defined to be the lowest of the ordinal categories. The following Tables show the posterior- predictive distribution of subject 10 and subject 12. These figures are calculated based on the simulated values of $\underline{\beta}^{(j)}, j = 1, 2, \ldots, m$. The sample size is also equal to 2,000 taken from the last 2,000 values of the entire simulated Markov chain.

Table 5.4. Results of Posterior-predictive distribution of subject 10 for given conditions (E198 experiment).

| Conditions | $\Pr(y^* =$ 'Bad' $\mid$ Cond.) | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0002 | 0.0163 | 0.0001 | 0.0010 |
| Cond 2 | 0.0003 | 0.0163 | 0.0001 | 0.0010 |
| Cond 3 | 0.0004 | 0.0169 | 0.0001 | 0.0011 |
| Cond 4 | 0.2459 | 0.3513 | 0.0718 | 0.5325 |

Table 5.4. (Continued) Results of Posterior-predictive distribution of subject 10 for given conditions (E198 experiment) .

| Conditions | $\Pr(y^* = \text{'Bad'} \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 5 | 0.0003 | 0.0152 | 0.0001 | 0.0008 |
| Cond 6 | 0.0035 | 0.0535 | 0.0006 | 0.0100 |
| Cond 7 | 0.1179 | 0.2792 | 0.0261 | 0.3254 |
| Cond 8 | 0.3799 | 0.3918 | 0.1308 | 0.6748 |

Table 5.5 Results of Posterior-predictive distribution of subject 12 for given conditions (E198 experiment).

| Conditions | $\Pr(y^* = \text{'Bad'} \mid \text{Cond.})$ | s.e. | $P_{0.025}$ | $P_{0.975}$ |
|---|---|---|---|---|
| Cond 1 | 0.0001 | 0.0107 | 0.0000 | 0.0004 |
| Cond 2 | 0.0002 | 0.0115 | 0.0001 | 0.0005 |
| Cond 3 | 0.0002 | 0.0122 | 0.0001 | 0.0006 |
| Cond 4 | 0.1544 | 0.3066 | 0.0281 | 0.3934 |
| Cond 5 | 0.0002 | 0.0120 | 0.0001 | 0.0005 |
| Cond 6 | 0.0019 | 0.0373 | 0.0003 | 0.0060 |
| Cond 7 | 0.0676 | 0.2214 | 0.0130 | 0.1968 |
| Cond 8 | 0.2564 | 0.3611 | 0.0571 | 0.5594 |

The figures in Table 5.4 and 5.5 reveal that given condition 4, 7 and 8, subject 10 and 12 are likely to give a response of having a bad connection. These figures are consistent with the results in Table 5.2, where parameters for these three conditions are highly significant. This is a straightforward calculation when conducting a Bayesian statistical analysis. As far as the telecommunication engineers are concerned, this is an important piece of information .

### 5.4.2 Model Checking Using Latent Residuals

Finally for model checking, we use again the method of latent residuals introduced by Albert and Chib (1995) in the continuation-ratio logit model. That is, if equation 5.3 is the correct model for the data,

$$\lambda_{ij} = \underline{x}'_{ij}\underline{\beta} + \varepsilon_{ij} \tag{5.9}$$

and $\lambda_{ij}$ are the latent variables for the h category, $\alpha_{h-1} < \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} < \alpha_h$ where $\alpha_{h-1}$ and $\alpha_h$ are the corresponding cut-points. As mentioned before, the latent variable $\lambda_{ij}$ follows a logistic distribution with mean $\underline{x}'_{ij}\underline{\beta}$ and unit variance. $\lambda_{ij}$ can be generated directly from a logistic distribution for given values of $\underline{\beta}^{(t)}$ and $\underline{b_i}^{(t)}$ at the $t$th stage of iteration of our Gibbs sampler scheme. Therefore, latent residuals $\varepsilon_{ij}$ are defined as

$$\varepsilon_{ij} = \lambda_{ij} - \underline{x}'_{ij}\underline{\beta} \tag{5.10}$$

Latent residuals $\varepsilon_{ij}$ are a priori random samples from a standard logistic distribution. In our studies, a sample size of 10,000 latent residuals have been generated from the model. To test whether or not the generated latent residuals follow a standard logistic distribution, a Kolmogorov-Smirnov test is used. In conducting the test a random sample of size 30 is sampled from the 10,000 generated latent residuals each time. We obtain the necessary $D_n$ test statistic as well as $\Pr(D \leq D_n)$. The test is repeated respectively for 50, 100 and 1000 times. Table 5.6 shows the overall results.

Table 5.6. Results of K-S Test of latent residuals

Continuation-ratio model (E198 Experiment)

| Sample size (n) | No. of runs (m) | Mean $D_n$ | Mean $\Pr(D \leq D_n)$ |
|---|---|---|---|
| 30 | 50 | 0.1653 | 0.4682 |
| 30 | 100 | 0.1738 | 0.4325 |
| 30 | 1000 | 0.1802 | 0.4190 |

The following histograms also show the distribution of the mean latent residuals and the plot of the mean latent residuals against each observation. There are altogether 256 mean latent residuals for the data. The latent residuals are obtained at each simulation run. The last 2,000 values of the entire 10,000 are used for plotting the histograms.

Figure 5.7: Histogram of the mean latent residuals superimposed by a standard logistic curve, n=256

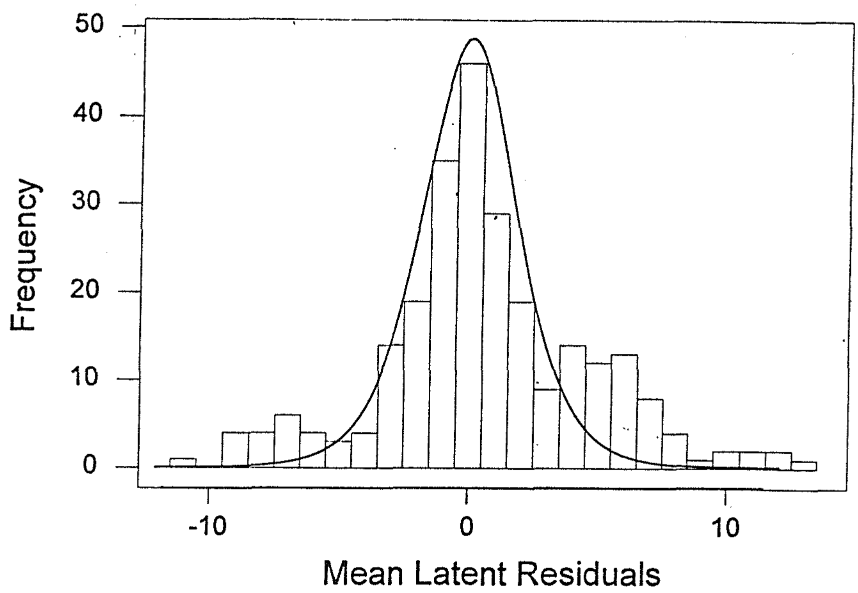
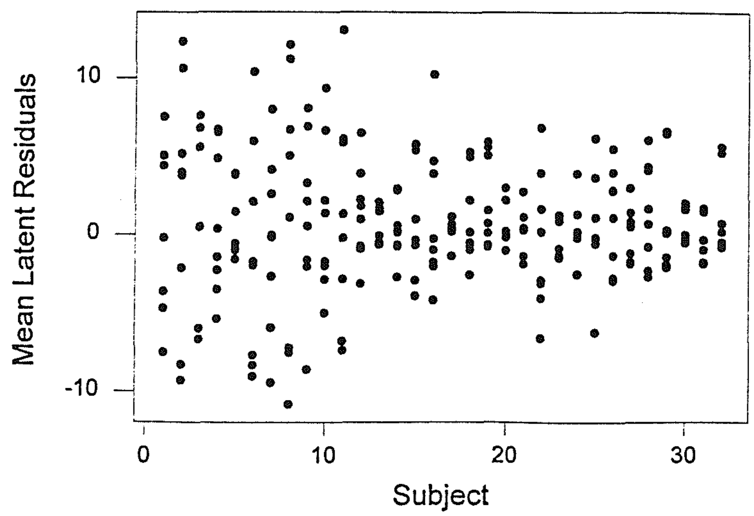



Figure 5.8: Plots of mean latent residuals against subjects



As we can see from Figure 5.7, the histogram is in close agreement with the superimposed standard logistic curve and there is no particular pattern exhibited in the scatter plot of mean latent residuals against the subjects.

## 5.5 Conclusions

In this chapter, we have only presented the estimation results of continuation-ratio model for the E198 data using the expanded likelihood. We assumed common parameters for each continuation ratio, apart from the intercept (cutpoint). This model could be extended to fitting a model for each level so there are four separate models. However, our experience indicates that after converting the data into binary form, that is, letting $Y_{ij} = 1$ if $Y_{ij} = h$ and $Y_{ij} = 0$ if $Y_{ij} > h$, the ARS estimation procedure failed to converge in each of the four cases due to the sparseness of "1" for each subject. Similar situations are encountered when we use the BUGS program for the same data set.

The interpretation of the continuation-ratio model is different from the model for strictly binary data in chapter 2. However, the estimation technique is the same. The proposed continuation-ratio model is a model which, in general, models any response of an ordered categorical nature. This implies that the model would be suitable to cope with the many other ordered categorical scales used in historical experiments. The traditional British Telecom method of analysis for opinion score responses has been used to perform an analysis-of-variance on the numerical scores assigned to the categories ( 0 to 4 ). One of the assumptions underlying the analysis-of-variance procedure is that the response variable follows a Normal distribution. The opinion score is constrained to one of five values, i.e., it is a discrete response rather than a continuous one. Approximating a discrete response with five values by a Normal curve is rather a crude approach. Also the scores attributed to different categories are arbitrary. The two considerations have the consequence of inefficient estimation of parameters by the standard analysis-of-variance approach.

Wolfe (1996) also used continuation-ratio model to fit the BT experimental data. He found that the residual sum of squares of the continuation-ratio models are in general smaller than the residual sum of squares as minimised in the analysis-of-variance approach. Thus the continuation-ratio logit model fitted the data better than the analysis-of-variance approach. The proposed residual sum of squares (Wolfe, 1996) for the continuation-ratio logit model $(RSS_{CR})$ is calculated as follows:

$$RSS_{CR} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5.11}$$

with $\hat{y}_i$ given by

$$\hat{y}_i = \sum_{j=0}^{r-1} j\hat{\pi}_{ij} \tag{5.12}$$

The fitted value $\hat{y}_i$ is calculated by multiplying the fitted probabilities from the continuation-ratio logit model $\hat{\pi}_{ij}$ by the scores $j = 0$ to 4 as used in the ANOVA model for the data, giving a fitted mean score for the continuation-ratio logit model.

The Bayesian method of modelling the continuation-ratio logit model using Markov chain Monte Carlo (MCMC) technique provides a good alternative way to model ordinal data. However, it is difficult to compare directly model fits via residual sums of squares. We use the concept of " latent residuals ". By inspecting the empirical posterior distribution of the mean latent residuals in Figure 5.7, we are confident that the model satisfies our basic assumption. That is, the distribution of the model error is distributed as standard logistic distribution. Also by inspecting the scatter plot of mean latent residuals against each subject, we note that no particular pattern is found in this plot. In the Bayesian approach, one can identify immediately those parameters which have significant effects by conveniently inspecting whether zero value is contained in the $(1 - \alpha) \times 100\%$ probability intervals. 95% probability intervals are adopted here.

Another advantage of Bayesian modelling is that we are able to compute the predicted probabilities of future events for given generated parameter values. When one uses Markov chain Monte Carlo (MCMC) technique for parameter estimation, one can also obtain the $(1-\alpha) \times 100\%$ probability intervals for the predicted probabilities of future events. More information about model prediction is incorporated naturally in the Bayesian approach. The model parameters and future value of the observations are random variables in the full probability model under discussion.

# Chapter 6

# Random Effects Cumulative Probit Model with Serial Correlation

## 6.1 Introduction

The purpose of this chapter is to consider models for ordinal observations with serial correlation. There has been a large amount of published work in recent years on the topic of repeated measures. Most of the published work has used a frequentist approach. Therefore it would be worthwhile to model these data using a Bayesian approach. Markov chain Monte Carlo method plays an indispensable role in model parameter estimation.

Suppose $Y_{it}$, $(Y_{it} \in \Re)$, $i = 1, 2, ..., n$ (subject), $t = 1, 2, ..., T_i$ ($T$ observations for each subject $i$), is a set of longitudinal data. To model these data collected over time for each member of a group of experimental units, one must recognize the possibility of correlation between serial observations on the same experimental unit. Several authors such as Potthoff and Roy (1964), Rao (1965, 1967) and Grizzle and Allen (1969) have analyzed balanced and complete longitudinal data using multivariate analysis of ANOVA models. However longitudinal data are often, in practice, unbalanced or incomplete, that is all individuals are not observed at the same number of time points or with the same design matrix. Chi and Reinsel (1989) consider the following model for longitudinal data that contain both individual random components and within-individual errors that follows an (autoregressive) AR(1) time series process. Their model for individual $i$ is

$$\mathbf{y_i} = \mathbf{X}_i\underline{\beta} + \mathbf{Z}_i\underline{b}_i + \underline{u}_i, \qquad i = 1, 2, .., n \tag{6.1}$$

where $\mathbf{y_i}$ is a $T_i \times 1$ vector of observations, $\mathbf{X}_i$ is the $T_i \times p$ design matrix for the mean vector of individual i, $\underline{\beta}$ is the p $\times$ 1 population fixed effect parameter vector, $\mathbf{Z}_i$ is the $T_i \times q$ design matrix for the random effects of individual i, $\underline{b}_i$ is a $q \times 1$ vector of unobservable random effects assumed to be sampled from a multivariate normal distribution with mean 0 and $q \times q$ covariance matrix $\Gamma$ and $\underline{u}_i$ is the $T_i \times 1$ vector of within-individual errors whose components are assumed to follow the AR(1) model,

$$u_{i,t} = \phi u_{i,t-1} + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma^2) \tag{6.2}$$

$\phi$ is the coefficient of the AR(1) process. For individual $i$, it is assumed that observations are taken at integer time points, $t_{i,1},...,t_{i,T_i}$, which are not necessarily consecutive, so the " missing data" situation is accommodated. Let $\sigma^2\Omega_t$ denote the covariance of $\underline{u}_i$, so

$$Cov(\mathbf{y_i}) = \mathbf{Z_i}\mathbf{\Gamma}\mathbf{Z_i'} + \sigma^2\mathbf{\Omega_t} \tag{6.3}$$

One may refer to the article by Chi and Reinsel (1989) for more details.

## 6.2 Modelling Longitudinal Ordinal Response in a Bayesian Perspective

If the longitudinal data $Y_i$ for individual $i$ is measured on an ordinal scale such as the BT ordinal response data in Chapter 3, then the above model will be invalid. Instead, we shall model the cumulative probability of the response category up to and including category $h$. A suitable link function should be used. Let $\gamma_{it,h} = \Pr(Y_{it} \leq h)$. Then

$$\begin{aligned} \text{Link}\{\Pr(Y_{it} \leq h)\} &= \text{Link}\{\gamma_{it,h}\} \qquad i = 1, ..., n, \; t = 1, ..., T_i \qquad (6.4) \\ &= \mathbf{X}_i\underline{\beta} + \mathbf{Z}_i\underline{b}_i + \underline{u}_i. \end{aligned}$$

$\mathbf{X}_i, \mathbf{Z}_i, \underline{\beta}, \underline{b}_i$ and $\underline{u}_i$, are the same as defined in section 1 of this chapter.

### 6.2.1 Random Effects Cumulative Probit Model with Serial Correlation

To answer the above question, we now propose a Bayesian approach to model data of this nature. This model is a random effects cumulative probit model with serial correlation. This model is also based on the concept of " latent variable", but instead of using the logit link, we use the probit link function. A random effects cumulative probit model with serial correlation would be of the form

$$\text{Probit}\,\{\gamma_{it,h}\} = \alpha_h - (\underline{x}'_{it}\underline{\beta} + \underline{z}'_{it}\underline{b}_i + e_{it}), \qquad (6.5)$$

where $e_{it} = \phi_{i1}e_{i,t-1} + a_{it}$ is assumed to be an autoregressive process of order one ( AR(1) ) and $\phi_{i1}$ is the AR(1) process parameter, $i = 1, 2, ..., n, \;\; t = 1, 2, ..., k, \;\; h = 1, 2, ...r$. $a_{it}$ are assumed to be Gaussian innovations with mean 0 and variance $\sigma^2$. Equation 6.5 can be rewritten as

$$\text{Probit}\,\{\gamma_{it,h}\} = \alpha_h - (\underline{x}'_{it}\underline{\beta} + \underline{z}'_{it}\underline{b}_i + \phi_{i1}e_{i,t-1} + a_{it}), \qquad (6.6)$$

As in chapter 3, it is assumed that there is a continuous 'latent' variable $\lambda_{it}$. In this case $\lambda_{it}$ is assumed to follow a Gaussian distribution with mean $\underline{x}'_{it}\underline{\beta} + \underline{z}'_{it}\underline{b}_i$ and variance $\sigma^2$. The observed ordinal data $Y_{it} = h, h = 1, 2, ..., r$ whenever $\alpha_{h-1} < \lambda_{it} \leq \alpha_h$.

Now for $t \geq 2$

$$a_{it} = e_{i,t} - \phi_{i1}e_{i,t-1}$$

$$= (\lambda_{it} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i) - \phi_{i1}(\lambda_{i,t-1} - \underline{x}'_{i,t-1}\underline{\beta} - \underline{z}'_{i,t-1}\underline{b}_i)$$
$$= \left[(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) - (\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\right] \quad (6.7)$$

$\underline{\beta}$ is the regression coefficient vector (p-dimension) in the model and $\underline{b}_i$ is the (q-dimension with q < p) coefficient vector for the random component. $\underline{b}_i$ is also assumed to follow a multivariate normal distribution with mean 0 and variance-covariance matrix $D$. In next section we shall present the Bayesian approach to formulate the model so that we can employ the Gibbs sampler to estimate various parameters in the model, such as $\underline{\beta}$, Var($D$), $\alpha_h$ and $\phi_{i1}$.

### 6.2.2 Bayesian Model Formulation

It is assumed that $a_{it}$'s are Gaussian innovations with mean 0 and variance $\sigma^2$. Let $\sigma^2 = \frac{1}{\tau}$, $\tau > 0$. $a_{it}$'s are *i. i. d.* variates. Then the likelihood function $L(\phi_{i1}, \underline{\beta}, \underline{b}_i, \tau, D \mid y_{ij})$ is given by

$$\begin{aligned} L \;=\; & \prod_{i=1}^{n}\prod_{t=2}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \le \alpha_h)\right] \times \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\ & -(\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\}^2 \\ & \times \mid D \mid^{\frac{1}{2}} \exp(-\frac{1}{2}\underline{b}_i{}'D^{-1}\underline{b}_i) \end{aligned} \quad (6.8)$$

If a non-informative prior distribution is used, that is,

$$P(\phi_{i1}, \underline{\beta}, \underline{b}_i, \tau, D) \propto \frac{1}{\tau}, \quad (6.9)$$

Then the joint posterior distribution is given by

$$\begin{aligned} P(\underline{\beta}, \underline{b}_i, D \mid y_{ij}) \;=\; & \frac{1}{\tau}\prod_{i=1}^{n}\prod_{t=2}^{k}\left[\sum_{h=1}^{r} \mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \le \alpha_h)\right] \times \\ & \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\ & -(\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\}^2 \\ & \times \mid D \mid^{\frac{1}{2}} \exp(-\frac{1}{2}\underline{b}_i{}^T D^{-1}\underline{b}_i) \end{aligned} \quad (6.10)$$

## 6.3 Conditional Posterior Distributions

The conditional posterior distribution of $\underline{\beta}$, denoted as $\left[\underline{\beta} \mid \cdot\right]$, is given by ( up to proportionality),

$$\begin{aligned}\left[\underline{\beta} \mid \cdot\right] &= \frac{\Pi(\underline{\alpha}, \lambda_{it}, \underline{\beta}, \underline{b}_i, D, \phi_{i1} \mid y_{ij})}{\int \cdots \int \Pi(\underline{\alpha}, \lambda_{it}, \underline{\beta}, \underline{b}_i, D, \phi_{i1} \mid y_{ij}) d\beta_0 d\beta_1 ... d\beta_{p-1}} \\ &= c_1 \prod_{i=1}^{n} \prod_{t=2}^{k} \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \leq \alpha_h)\right] \times \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\ &\quad -(\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\}^2, \end{aligned} \quad (6.11)$$

where $c_1$ is a constant. That is

$$\left[\underline{\beta} \mid \cdot\right] = \left[\underline{\beta} \mid \underline{\alpha}, \lambda_{it}, \underline{b}_i, \phi_{i1}, y_{ij}\right]$$

For fixed $i$, the conditional marginal posterior distribution of $\underline{b}_i$, denoted as $[\underline{b}_i \mid \cdot]$, is given by ( up to proportionality),

$$\begin{aligned}[\underline{b}_i \mid \cdot] &= \frac{\Pi(\underline{\alpha}, \lambda_{it}, \underline{\beta}, \underline{b}_i, D, \phi_{i1} \mid y_{ij})}{\int \cdots \int \Pi(\underline{\alpha}, \lambda_{it}, \underline{\beta}, \underline{b}_i, D, \phi_{i1} \mid y_{ij}) db_0 db_1 ... db_q} \\ &= c_2 \prod_{t=2}^{k} \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \leq \alpha_h)\right] \times \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\ &\quad -(\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\}^2 \times \exp(-\frac{1}{2}\underline{b}_i{}' D^{-1}\underline{b}_i) \end{aligned} \quad (6.12)$$

where $c_2$ is a constant. That is

$$[\underline{b}_i \mid \cdot] = \left[\underline{b}_i \mid \underline{\alpha}, \lambda_{it}, \underline{\beta}, \phi_{i1}, y_{ij}\right]$$

For fixed $i$ and $t$,

$$\begin{aligned}[\lambda_{it} \mid \cdot] &= c_3 \times \left[\sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \leq \alpha_h)\right] \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\ &\quad -(\underline{x}'_{it}\underline{\beta} + \phi_{i1}\underline{x}'_{i,t-1}\underline{\beta}) - (\underline{z}'_{it}\underline{b}_i + \phi_{i1}\underline{z}'_{i,t-1}\underline{b}_i)\}^2. \end{aligned} \quad (6.13)$$

That is,

$$[\lambda_{it} \mid \cdot] = \left[\lambda_{it} \mid \underline{\alpha}, \underline{\beta}, \underline{b}_i, \phi_{i1}\right].$$

$c_3$ is also a constant. It can be shown further that

$$[D \mid \cdot] \;=\; [D \mid \underline{b}_i]$$
$$[\phi_{i1} \mid \cdot] \;=\; \left[\phi_{i1} \mid \underline{\alpha}, \lambda_{it}, \underline{\beta}, \underline{b}_i, D, y_{ij}\right]$$

Finally the full conditional distribution for each of the cutpoints $\alpha_h, h = 2, 3, ..., r-1$ ($\alpha_1$ is being fixed at 0) is again given by, (upto proportionality ),

$$\begin{aligned} [\alpha_h \mid \cdot] \quad & \propto \prod_{i=1}^{n} \prod_{j=1}^{k} [I(y_{it} = h) I(\alpha_{h-1} < \lambda_{it} \leq \alpha_h) \\ & + [I(y_{it} = h+1) I(\alpha_h \leq \lambda_{it} < \alpha_{h+1})] , \end{aligned} \tag{6.14}$$

## 6.4 Adaptive Rejection Sampling

As in the previous chapters, the adaptive rejection sampling method is employed to generate all the parameters in this probit model with serial correlation. It is not difficult to prove that the conditional posterior distributions, $h(\underline{\beta}), h(\underline{b}_i), h(\lambda_{it})$ and $h(\phi_i)$ are all log-concave. The proofs are shown as follows. To estimate $D$, again the Odell-Feiveson (1966) algorithm is used.

## 6.5 Proofs of Log-concavity Conditions

To check whether the conditional marginal posterior distribution of $\underline{\beta}$ satisfies the log-concavity condition, we first consider equation 6.11 and let

$$\begin{aligned} h(\underline{\beta}) = \quad & c_1 \prod_{i=1}^{n} \prod_{t=2}^{k} \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h) \mathrm{I}(\alpha_{h-1} < \lambda_{it} \leq \alpha_h) \right] \times \frac{\tau}{\sqrt{2\pi}} \exp -\frac{\tau}{2} \{ (\lambda_{it} - \phi_{i1} \lambda_{i,t-1}) \\ & - (\underline{x}_{it}' \underline{\beta} + \phi_{i1} \underline{x}_{i,t-1}' \underline{\beta}) - (\underline{z}_{it}' \underline{b}_i + \phi_{i1} \underline{z}_{i,t-1}' \underline{b}_i) \}^2 \end{aligned} \tag{6.15}$$

If natural logarithm is taken on both sides of equation 6.15, then equation 6.15 becomes

$$\ln h(\underline{\beta}) = \quad c + \sum_{i=1}^{n} \sum_{t=2}^{k} \ln \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h) \mathrm{I}(\alpha_{h-1} < \lambda_{it} \leq \alpha_h) \right]$$

$$+n(k-1)\ln\tau - \frac{\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{k}\{(\lambda_{it}-\phi_{i1}\lambda_{i,t-1})$$
$$-(\underline{x}_{it}'\underline{\beta}+\phi_{i1}\underline{x}_{i,t-1}'\underline{\beta})-(\underline{z}_{it}'\underline{b}_i+\phi_{i1}\underline{z}_{i,t-1}'\underline{b}_i)\}^2, \qquad (6.16)$$

where $c$ is a constant. Then

$$\frac{\partial \ln h(\underline{\beta})}{\partial \underline{\beta}} = -\tau\sum_{i=1}^{n}\sum_{t=2}^{k}(-\underline{x}_{it}'+\phi_{i1}\underline{x}_{i,t-1}')\{(\lambda_{it}-\phi_{i1}\lambda_{i,t-1})$$
$$-(\underline{x}_{it}'+\phi_{i1}\underline{x}_{i,t-1}')\underline{\beta}-(\underline{z}_{it}'+\phi_{i1}\underline{z}_{i,t-1}')\underline{b}_i\} \qquad (6.17)$$

$$\begin{aligned}\frac{\partial^2 \ln h(\underline{\beta})}{\partial \underline{\beta}\partial \underline{\beta}'} &= -\tau\sum_{i=1}^{n}\sum_{t=2}^{k}(\underline{x}_{it}\underline{x}_{it}'+2\phi_{i1}\underline{x}_{it}\underline{x}_{i,t-1}'+{\phi_{i1}}^2\underline{x}_{i,t-1}\underline{x}_{i,t-1}') \\ &= -\tau\sum_{i=1}^{n}\sum_{t=2}^{k}\underline{u}_{it}\underline{u}_{it}' \\ &= -\tau\sum_{i=1}^{n}\sum_{t=2}^{k}A_{it}, \end{aligned} \qquad (6.18)$$

where $\underline{u}_{it}=\underline{x}_{it}+\phi_{i1}\underline{x}_{i,t-1}$ and $A_{it}=\underline{u}_{it}\underline{u}_{it}'$ is a symmetric matrix for each $i$ and $t$. Matrix $A_{it}$ is necessarily positive-semi definite. Thus $h(\underline{\beta})$ is log-concave. Similarly, for fixed $i$, we can prove that

$$\begin{aligned}\frac{\partial^2 \ln h(\underline{b}_i)}{\partial \underline{b}_i\partial \underline{b}_i'} &= -\tau\sum_{t=2}^{k}(\underline{z}_{it}\underline{z}_{it}'+2\phi_{i1}\underline{z}_{it}\underline{z}_{i,t-1}'+{\phi_{i1}}^2\underline{z}_{i,t-1}\underline{z}_{i,t-1}')-D^{-1} \\ &= -\tau\sum_{t=2}^{k}\underline{v}_{it}\underline{v}_{it}'-D^{-1} \\ &= -\tau\sum_{t=2}^{k}B_{it}-D^{-1}, \end{aligned} \qquad (6.19)$$

where $\underline{v}_{it}=\underline{z}_{it}+\phi_{i1}\underline{z}_{i,t-1}$ and $B_{it}=\underline{v}_{it}\underline{v}_{it}'$ and $D^{-1}$ are q $\times$ q symmetric matrices. Matrix $B_{it}$ and $D^{-1}$ are also positive-semi definite. Therefore $h(\underline{b}_i)$ is log-concave.

To show whether the conditional posterior distribution of the coefficient, $\phi_{i1}$, of the AR(1) process is log-concave, consider

$$\begin{aligned}h(\phi_{i1}) &= \frac{\Pi(\underline{\alpha},\lambda_{it},\underline{\beta},\underline{b}_i,D,\phi_{i1}\mid y_{ij})}{\int\Pi(\underline{\alpha},\lambda_{it},\underline{\beta},\underline{b}_i,D,\phi_{i1}\mid y_{ij})d\phi_{i1}} \\ &= h(\phi_{i1}\mid \underline{\alpha},\lambda_{it},\underline{\beta},\underline{b}_i,D,y_{ij}) \\ &= c_4\prod_{t=2}^{k}\left[\sum_{h=1}^{r}\mathrm{I}(y_{ij}=h)\mathrm{I}(\alpha_{h-1}<\lambda_{it}\le\alpha_h)\right]\times\frac{\tau}{\sqrt{2\pi}}\exp-\frac{\tau}{2}\{(\lambda_{it}-\phi_{i1}\lambda_{i,t-1}) \\ &\quad -(\underline{x}_{it}'\underline{\beta}+\phi_{i1}\underline{x}_{i,t-1}'\underline{\beta})-(\underline{z}_{it}'\underline{b}_i+\phi_{i1}\underline{z}_{i,t-1}'\underline{b}_i)\}^2, \end{aligned} \qquad (6.20)$$

where $c_4$ is a constant. Taking natural logarithm on both sides of 6.20, we have

$$\begin{aligned}
\ln h(\phi_{i1}) \quad &= \ln c_4 + \sum_{t=2}^{k} \ln \left[ \sum_{h=1}^{r} \mathrm{I}(y_{ij} = h)\mathrm{I}(\alpha_{h-1} < \lambda_{it} \le \alpha_h) \right] \\
&-\frac{\tau}{2}\sum_{t=2}^{k}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) \\
&-(\underline{x}_{it}'\underline{\beta} + \phi_{i1}\underline{x}_{i,t-1}'\underline{\beta}) - (\underline{z}_{it}'\underline{b}_i + \phi_{i1}\underline{z}_{i,t-1}'\underline{b}_i)\}^2. \qquad (6.21)
\end{aligned}$$

Then

$$\begin{aligned}
\frac{\partial \ln h(\phi_{i1})}{\partial \phi_{i1}} = \quad & +\tau\sum_{t=2}^{k}\{(\lambda_{it} - \phi_{i1}\lambda_{i,t-1}) - (\underline{x}_{it} + \phi_{i1}\underline{x}_{i,t-1})\underline{\beta} \\
& -(\underline{z}_{it}' + \phi_{i1}\underline{z}_{i,t-1}')\underline{b}_i\}(\lambda_{i,t-1} - \underline{x}_{i,t-1}'\underline{\beta} - \underline{z}_{i,t-1}'\underline{b}_i) \qquad (6.22)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ln h(\phi_{i1})}{\partial \phi_{i1}^2} \quad &= -\tau\sum_{t=2}^{k}(\lambda_{i,t-1} - \underline{x}_{i,t-1}'\underline{\beta} - \underline{z}_{i,t-1}'\underline{b}_i)^2 \\
\frac{\partial^2 \ln h(\phi_{i1})}{\partial \phi_{i1}^2} \quad &< 0;
\end{aligned}$$

since $\tau = \frac{1}{\sigma^2} > 0$. So the conditional posterior distribution of $\phi_{i1}$ is log-concave. Lastly, for the latent continuous variable $\lambda_{it}$, it is not difficult to show that the conditional posterior distribution $h(\lambda_{it})$ is log-concave where for fixed $i$ and $t$,

$$\frac{\partial \ln h(\lambda_{it})}{\partial \lambda_{it}} = -\tau\{(\lambda_{it} - \underline{x}_{it}'\underline{\beta} - \underline{z}_{i,t-1}'\underline{b}_i) - \phi_{i1}(\lambda_{i,t-1} - \underline{x}_{i,t-1}'\underline{\beta} - \underline{z}_{i,t-1}'\underline{b}_i)\}$$

$$\frac{\partial^2 \ln h(\lambda_{it})}{\partial^2 \lambda_{it}} = -\tau < 0,$$

since $\frac{1}{\tau} = \sigma^2 > 0$.

The results in this section also confirmed the results stated by Wedderburn (1976) that the generalized linear models with many link functions , such as normal, logit, probit, Poisson and complementary log-log links, in the exponential family satisfy the concavity condition. In the next section the ARS method is used in a Gibbs sampling scheme to generate all the model parameters.

## 6.6 Simulation Studies of Probit Model with Serial Correlation

Having proved the log-concavity conditions for the conditional posterior distributions of the model parameters, simulation work is presented in this section. The main purpose is to demonstrate the viability of the proposed probit model with serial correlation for ordered categorical data. As in chapter 3 and 4, the following cumulative probability model with serial correlation using a probit link function is considered:

$$\begin{aligned} \text{probit}P(Y_{it} \leq h \mid \underline{b}_i) = \quad & \alpha_h - (\beta_0 + \beta_1 T + \beta_2 X_i + \beta_3 T X_i + b_{0i} + b_{1i} T) \\ & + e_t + a_t \end{aligned} \tag{6.23}$$

where $X_i = 0$ for half of the population and 1 for the remainder and $T = -3, -2, -1, 0, 1, 2, 3$. The fixed effects coefficients were set at $\beta_0 = +3, \beta_1 = -2.5, \beta_2 = +2.5$ and $\beta_3 = -1.0$. The random effects distribution is simulated with

$$\text{Var}(\underline{b}_i) = D = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.001 \end{bmatrix} \tag{6.24}$$

and $e_t = \phi_{1i} e_{t-1} + a_t$ is an order one autoregressive time series (AR(1)). $a_t$ is a Gaussian innovation with mean zero and variance $\frac{1}{\tau}$, $\tau = \frac{1}{\sigma^2} > 0$. In this simulation, $\phi_{1i}$, the AR(1) coefficient is set equal to +0.5. The AR(1) time series with $\phi_{1i}$ is generated using the NAG-FORTRAN subroutines library.

The simulation work is carried out using non-informative prior and $\tau = 1$. However, past experience of other researchers such as Palmer and Broemeling (1993) and Chib (1993) indicated that if a Gibbs sampling approach is adopted, there are certain risks in using a non-informative prior distribution to estimate the regression coefficients and autocorrelation parameter. Palmer and Petit (1995) gave an numerical example using simulated data to show that the final estimates were far away from the original values of the parameters. The problem will get worse if the autocorrelation parameter is close to +1.0 or -1.0. A unit root nonstationary phenomenon will occur which leads to a so-called 'black-hole' problem in the estimation procedure. Chib (1993) has suggested that an improper prior can be used to implement a solution using Gibbs sampling when moderate-to-high positive autocorrelation was presented. However, Palmer and Petit (1995) recommended that by using a fairly 'vague' proper prior, the difficulty can be resolved. They also gave an numerical example with simulated data to support their recommendation.

### 6.6.1 Results of Simulation

Similar difficulties are also encountered in the simulation work here if a non-informative prior with $\tau = 1$ is used. The final results are shown in Table 6.1. The overall results are based on one typical run using one simulated data set. A sample size of n=11,000 is generated and we discard the first 1,000 values as burn-in.

Table 6.1. Results of Simulation Studies for Probit Model

using non-informative prior

| Parameter | True value | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| $\beta_0$ | 3.0 | -0.8407 | -0.8395 | 1.5078 | -5.2536 | 4.3467 |
| $\beta_1$ | -2.50 | -4.5156 | -4.5108 | 1.3911 | -7.6542 | -0.5467 |
| $\beta_2$ | +2.50 | 4.3811 | 4.4431 | 3.4462 | 1.1702 | 8.2156 |
| $\beta_3$ | -1.00 | -0.7062 | -0.7045 | -1.6210 | -3.0156 | 4.5891 |
| Var($b_0$) | +1.00 | 4.8851 | 4.1509 | 1.8272 | 0.9263 | 9.5663 |
| $\phi_1$ | +0.5 | 0.2324 | 0.2311 | 0.1164 | 0.0034 | 0.6651 |

The final estimates of the parameters using a non-informative prior are not good. Only $\beta_3$ is closer to the true value. The worst estimate is the random effects component and there is about 50% underestimate for the autocorrelation coefficient. The value for PSRF is about 2.7815 which indicates the whole Gibbs sampling scheme does not converge well. Therefore there is reason to believe that a vague proper prior should be used in order to obtain a better estimate from the ARS-Gibbs sampling scheme. Here a multivariate normal distribution for the regression coefficient vector $\underline{\beta}$ is chosen as the 'vague' proper prior. That is, the density of the prior distribution $P(\underline{\beta})$ is given by,

$$P(\underline{\beta}) = \frac{1}{(2\pi)^{\frac{p}{2}} \mid V \mid^{\frac{p}{2}}} \exp\{(\underline{\beta} - \underline{\mu}_0)' V^{-1} (\underline{\beta} - \underline{\mu}_0), \qquad (6.25)$$

where $V = \text{Diag}(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{p-1}^2)$ and $\sigma_i^2 = Var(\beta_i)$.

In this simulation study, $\sigma_0^2$ is set equal to 0.5 for the mean location parameter $\beta_0$ and $\sigma_i^2 = 1.0$ for all other parameters of $\underline{\beta}$ (Palmer and Petit, 1995) .

Since this is a simulation work and all the true parameters are known, the mean vector $\underline{\mu}_0$ is set at the values which generated the data for simulation. It is not difficult to prove that the conditional posterior distributions also satisfy the log-concavity conditions if a multivariate normal prior distribution is chosen for $\underline{\beta}$. The results of the ARS-Gibbs sampling scheme is shown in Table 6.2.

Table 6.2. Results of Simulation Studies for Probit Model

using vague proper prior

| Parameter | True value | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| $\beta_0$ | 3.0 | 3.0040 | 3.0043 | 0.1633 | 2.5994 | 3.3952 |
| $\beta_1$ | -2.50 | -2.5157 | -2.5151 | 0.0505 | -2.8394 | -2.2086 |
| $\beta_2$ | +2.50 | 2.4901 | 2.4885 | 0.0651 | 2.2302 | 4.0685 |
| $\beta_3$ | -1.00 | -0.9677 | -0.9829 | 0.1008 | -2.1657 | -0.4906 |
| Var($b_0$) | +1.00 | 1.1998 | 1.1273 | 1.2153 | 0.0440 | 4.1998 |
| $\phi_1$ | +0.5 | 0.4165 | 0.4203 | 0.0991 | -0.1558 | 0.7221 |

The results in Table 6.2 show a remarkable improvement over those in Table 6.1. All the estimates are close to the original parameters except the AR(1) coefficient $\phi_1$ which is a slight underestimate. This is largely due to the short length of the series for each subject $i$. However, in real practical situations, this is often the case. Long series of observations for a single subject are rare.

## 6.7 Applications of Probit Model with Serial Correlation

The problem of choosing *probit* or *logit* models for analyzing practical data has been discussed in previous chapters. The main conclusion is that in many situations both models would fit well. The main difference is the measurement scales of the model parameters. However if one wishes to fit a model to a set of repeated measure data which contains serial correlation within each individual subject under observation, then a probit model will be easier to handle than a logit model. The innovations in the probit model are assumed to follow the Gaussian distribution. It is not the case in cumulative logit model. If a time series model is driven by non-Gaussian innovations, it is difficult to conduct statistical inference on the model parameters.

As far as applications of the random effects probit model with serial correlation is concerned, this type of model has wide applications in many field of studies. Like the cumulative logit model for ordinal data, one can use the random effects probit model with serial correlation to model ordinal data whenever the investigator suspects that the data might exhibit serial correlations. However, the observed ordinal data should be taken from approximately equally-spaced time periods. For the BT experimental data, it is not sure that the successive observations are obtained over equally-spaced time periods given the covariates. Some of their experiments are of limited durations and some are of unlimited duration. Therefore it would be inappropriate to use a probit model with serial correlation to model their data. The BT data may best be modelled by the cumulative logit model with carry-over effects. By carry-over effects, we mean that in an experiment, each of $s$ subjects receives a sequence of $t$ treatments in a series of $p$ consecutive time periods. A response is measured at the end of each period. In planning such an experiment, and analyzing the results, the extent to which each treatment influences the responses in periods following the period of application should be considered. The existence of such $carry-over$ or *residual* effects depends on the particular problem under investigation. For example, in some pharmacokinetic studies it may be possible to employ washout periods of sufficient length to allow dissipation of any residual treatment period effects before the next treatment period begins; see for example Senn (1993). In other areas, an assumption of negligible carry-over effects at the design stage may be inappropriate, particularly if psychological carry-over is likely; see Jones and Kenward (1989). Also it is generally assumed that a carry-over effect from the current treatment period into the next period may occur, but effects persisting for more than one period are believed to be much smaller, and are therefore neglected. Effects persisting for only one period after the period of application are usually called first order carry-over effects.

The approach to modelling carry-over effect that is now introduced involves investigating an interaction between current response and previous response. This may be done by altering the formulation of the continuation-ratio model in equation 5.5 of Chapter 5, that is,

$$\log\left[\frac{\pi_{ij,h}}{1-\gamma_{ij,h}}\right] = \alpha_h - \underline{x}'_{ij}\underline{\beta}, \tag{6.26}$$

to

$$\log\left[\frac{\pi_{ij,hr}}{1-\gamma_{ij,hr}}\right] = \alpha_h + \alpha_{hr} - \underline{x}'_{ij}\underline{\beta} \tag{6.27}$$

where $r$ denotes the previous response. Fitting this model will give a matrix of parameters with elements $\alpha_{hr}$ and of dimension $(R-1)\times R$ where $R$ is the number of categories. The cut-point parameters $(\alpha_h)$ are also in the model. There will be $R-1$ redundant parameters among the $\alpha_{hr}$. A Bayesian approach to model a carry-over effect model in equation 6.27 using MCMC method is not difficult. However in the findings of Wolfe (1996), the suspected carry-over effects in the BT experimental data might well be explained by the subject scaling term in the model. It is possible that what has been attributed to carry-over of response could in fact be a subject-specific effect. Therefore we do not pursue further research on the carry-over effect model here.

## 6.8 Conclusion

Past experiences by other researchers (Palmer and Broemeling (1993), Palmer and Petit (1995), Chib (1993)) showed that it is risky to use non-informative prior for Bayesian parameter estimation in a model with serial correlation. Also there is difficulty in using Gibbs sampling to estimate the autocorrelation coefficient in a linear regression model when moderate-to-high positive autocorrelated errors are present. This phenomenon is also found in the probit model with serial correlation when a non-informative prior is used. Results showed by Palmer and Petit (1995) and also results from the simulation work of this chapter confirmed the fact that this difficulty may be resolved when a 'vague' proper prior distribution is used. The problem only occurs where there is an intercept in the model and if there is no intercept, as described by Palmer and Petit (1995), the problem does not arise.

The same problem could also occur when using an AR(p) model in the time series representation when using a non-informative prior if the sum $\sum_{i=1}^{p} \rho_i$ were close to 1. The distribution of the intercept would become singular. A so-called unit-root non-stationary problem would arise regardless of the dimension of the problem and the higher the order of the AR model, the more difficult this problem will be to solve.

In summary, it can very risky to rely on an improper prior as an automatic choice in estimation using Gibbs sampling. These findings have important implications in solving practical problems. Treviño-Villarreal (1999) has analyzed a large set of ordered categorical data using the TSPACK package. The results are obtained via maximum likelihood estimation. The observed data ($Y_{it}$) are the credit ratings of a certain country $i$ over a certain period $t$. The observations $Y_{it}$ for each country have a time dependence structure with very high correlations. The associated covariates, $X_{it}$, are a set of economic indicators such as annual economic growth rate, GNP, Balance of Trade, etc. In her thesis, she has not calculated the first order correlation coefficient of credit rating for each country over the observed time period. This is mainly because TSPACK package does not contain any program to calculate this AR(1) coefficient.

One needs either to write a special program within TSPACK package or resort to the methods developed in this chapter. The Bayesian way of modelling suggested in this chapter will be a very useful tool in modelling such data set. This model also has potentially wide applications in many fields of studies, such as in medical research, psychological research etc.

# Chapter 7

# Concluding remarks

The purpose of this chapter is to conclude some of the main points from the thesis and to point to further work. There has been a large amount of published work in recent years on the topic of repeated measures either in the frequentist or Bayesian approach. The ordinal responses of the BT experiments as well as the rating data of chapter 6 are longitudinal in nature. A discussion of how the approach taken in this thesis to modelling these responses relates to other methods of modelling repeated ordinal responses is worth mentioning here.

## 7.1 Conclusions

The main tools used in this thesis are primarily Bayesian methods with the application of Gibbs sampling. One main reason of using a Gibbs sampling approach to fitting the cumulative logit and probit model is because it can deal with random effects quite easily. The methodologies discussed in the thesis are versatile in their own right. Pan and Thompson (1998) pointed out that if a numerical integration scheme is used to estimate the random effects component; serious difficulty will be encountered if the dimensions of the integration are high. This situation will not happen if one uses the

Gibbs sampling scheme (ARS method) discussed in the thesis. Markov chain Monte Carlo (MCMC) methods make possible the use of flexible Bayesian models that would otherwise be computationally infeasible.

In chapter 2, we have discussed the use of ARS method to model longitudinal binary data using a Bayesian approach. Simulation work shows that the method works well even though it can be handled quite easily by a software developed by Biostatistics Unit of Cambridge University called "BUGS". But by writing our own code, we have confidence to solve more complicated problems such as the model for the BT binary data. The BT binary case requires more computational skills. Having successfully solved this problem, we are in a position to solve the problem stated in the article by Pan and Thompson (1998). The research will focus on the use of MCMC method to solve generalized linear mixed models. These models are widely used to model animal breeding experiments where either binary or binomial data with within-cluster correlation are concerned. Statistical inferences for generalized linear mixed models, however, are greatly hampered by the need for numerical integrations.

In chapter 3, a cumulative logit model for repeated ordinal responses is discussed. The Bayesian way of modelling cumulative logit model is presented. There are a large amount of work which have been published about this topics in recent years. The Bayesian approach with the use of MCMC method is now becoming more popular as a realistic way of statistical modelling. It has gained wide applications in many areas recently. It is also due to a great leap forward in computer technologies in the past decades. Computing nowadays is cheap. Bayesian statistical modelling relies heavily on statistical computations. So it is easy to do Bayesian statistics because of the wide availability of computers in academic environments.

In chapter 3, we presented a simulation study on modelling repeated ordinal responses using the ARS method. The results are satisfactory. Albert and Chib (1993) use the probit link model. We use the cumulative logit link model. In fact to model categorical data using either logit and probit will give similar results except in extreme cases. A defense for the use of cumulative logit link is given in chapter 2. We also use the method to analyze the BT experimental data with reasonable results. The empirical results show that our method can apply to the so-called Rasch models which are widely used in the field of Psychological studies. The ARS method can also be used to estimate the parameters in the Rasch model. It is our conjecture that cumulative logit model with random effects using ARS method can solve generalized linear mixed models with repeated ordinal responses; a rich research area waiting to be exploited.

It is suspected that the BT experimental data may be modelled with a so-called carry-over effect. A carry-over effects cumulative logit link model should be considered for further modelling. However, the finding by Wolfe (1996) suggested that it is possible what is being called a carry-over effect of response can alternatively be explained by the model with a subject-specific scaling effect. These considerations lead to the analysis in chapter 4 where in addition to a subject-specific location effect, a subject-specific scaling effect is investigated. The need for a subject-specific approach to longitudinal ordinal response data as suggested in chapter 4 has been recognized at least since Torgerson (1958). The model used to perform these subject-specific analyzes is also a cumulative logit model. The inclusion of a scaling term in the cumulative logit model is straightforward. Although not all the parameters in the new models satisfied the log-concave condition; it is a good effort to use the Metropolis scheme to generate the subject-specific scaling term in the model.

The discussion of scaling terms in this thesis makes it clear that there is more than one possible formulation of the cumulative logit model when a scaling term is included. Two formulations are considered. One is the form considered by McCullagh and the other is the Kijewski's form. Both forms have a place in the analyst's toolkit. Kijewski's form of a subject-specific scaling term implies that the dispersion of the underlying distributions are the same for all subjects but that the subjects differ in the way they interpret the scale.

As mentioned in chapter 2, the ARS method has its drawbacks. For large models the computational speed is relatively slow and the serial correlation is rather high for each output series. For future research, it is also our conjecture that the methodology proposed in chapter 4 can solve similar problems if the random components have more complicated structure, such as those in generalized linear mixed models. However the computational efforts will be considerable.

In chapter 5, the continuation-ratio model is discussed. Again a Gibbs sampler is used to estimate the model parameters. We employed the techniques used in estimating model parameters for binary data in chapter 2. Therefore technically modelling the continuation-ratio model posed no problem here. The ways that we interprete model parameter are more important. This is because continuation-ratio model is to model probability of category $j$ given that categories $j, j+1, \ldots, K$ have ocurred. We presented a simulated example and a real application on BT telephone experimental data.

Throughout chapter 3 to 5, we use ARS method in the Gibbs sampler scheme to generate model parameters. The link functions are logistic. In fact we can either use a logit or probit model. The final analyses are quite similar. The only differences are in the scale of estimation.

In chapter 6, a cumulative probit model with serial correlation is considered. This kind of model involves a time series representation for one component in the model. Therefore the use of probit link in this model will facilitate parameter estimation in the time series component. This is because other than probit link (the time series is driven by Gaussian innovations), estimation will become very difficult. Also we proved in this chapter that the log-concave condition is satisfied for any design matrix in this model. We can use ARS method to estimate the coefficients of the time series component. Results in simulation work show that the method work well. Finally our work can be summarized by the following list of models starting from model for binary data,

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \underline{x}'_{ij}\underline{\beta} + \underline{z}'_{ij}\underline{b}_i, \qquad \Pr(Y_{ij}=1)=p_{ij}$$

$$\log\left(\frac{\gamma_{ij,h}}{1-\gamma_{ij,h}}\right) = \alpha_h - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i, \qquad \Pr(Y_{ij}\le h)=\gamma_{ij,h}$$

$$\log\left(\frac{\gamma_{ij,h}}{1-\gamma_{ij,h}}\right) = \alpha_h e^{\tau_i} - \underline{x}'_{ij}\underline{\beta} - \underline{z}'_{ij}\underline{b}_i, \qquad \Pr(Y_{ij}\le h)=\gamma_{ij,h}$$

$$\Phi^{-1}(\gamma_{it,h}) = \alpha_h - \underline{x}'_{it}\underline{\beta} - \underline{z}'_{it}\underline{b}_i + \phi_{i1}e_{i,t-1}, \qquad \Pr(Y_{it}\le h)=\gamma_{it}$$

## 7.2 Suggestions for future work

The thesis so far only examines models for longitudinal binary and ordinal data in a Bayesian perspective. There is a wide scope for future research. Firstly, the present methodology can be further developed to cater multivariate logit and probit models. Chib and Greenberg (1998) present an article on a practical simulation-based Bayesian and non-Bayesian analysis of correlated binary data using the multivariate probit model. The multivariate logit model using ARS method would be worth examining further.

As mentioned earlier, our method can be extended to solve GLMM *Model II* problems suggested by Pan and Thompson (1998). GLMM *Model I* is simply GLM models with random effects. GLMM *Model II* is a more general case where several mutually independent random effects $b_1, b_2, ..., b_c$ are incorporated simultaneously into the linear predictor $\eta = X\beta + Z_1b_1 + \cdots + Z_cb_c$, where $b_l \sim N_{ql}(0, \Sigma_l)$ $(l = 1, ..., c)$. In GLMM *Model II*, the variance-covariance matrices $\Sigma_l$'s are not diagonal. That is, correlations may occur amongst the random effects components. This situation often found in many scientific investigations such as animal breeding experiments in biological sciences, where either binary or binomial data within-cluster correlation are commonly concerned (see Thompson, 1990). A Gibbs sampler using the ARS method is a good alternative to the quasi-monte carlo EM algorithm suggested by Pan and Thompson (1998). Their method often encounters computational difficulties when the dimension of integration of the likelihood function is high.

Lastly, it is worth mentioning that the computer programs used in this thesis are entirely written in FORTRAN running on Unix workstations. This is to link with the ARS source program supplied by Dr. Gilks of Cambridge University. The Medical Research Council Biostatistics Unit at Cambridge University developed a software called the "BUGS" which has been mentioned earlier. BUGS is a useful software for MCMC. However the early versions of the program could handle mostly just "toy" problem, and were fairly buggy (comments by Carlin, Kass, *el at.*, 1998). The most updated (1998) new window version of BUGS software package called "WinBUGS" has improved quite a lot. Therefore in the future we can use "WinBUGS" software on a PC platform to carry out our research in modelling ordinal data or to perform Bayesian statistical analysis when MCMC methodology is required.

# References

Agresti, J. H. (1996). *An Introduction to Categorical Data Analysis.* Wiley, New York.

Agresti, J. H. and Lang J. B. (1993). A Proportional Odds Model with Subject-Specific Effects for Repeated Ordered Categorical Responses. *Biometrika*, **80**, 527-534.

Aitchison, J. and Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, **29**, 813-828.

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669-679.

Albert, J. and Chib, S. (1995). Bayesian Residual Analysis for Binary Response Regression Models. *Biometrika*, **82**, 747-759.

Albert, J. and Chib, S. (1996). Bayesian probit modeling of binary repeated measures data with an application to a cross-over trial. *Bayesian Biostatistics Casebook* , Marcel Dekker.

Anderson, D. A. and Aitkin, M. (1985). Variance Component Models with Binary Responses:Interviewer Variablility. *Journal of the Royal Statistical Society*, **B**, **47**, 203-210.

Ashford J. R. (1959). An approach to the analysis of data for semi-quantal responses in biological assay. *Biometrics*, **15**, 573-581.

Booth J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society*, **B, 61**, 265-285.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.

Breslow, N. E. (1984). Extra-Poisson Variation in Log-Linear Models. *Applied Statistics*, **33**, 38-44.

CCITT (1992). *Handbook on Telephonometry*, Geneva.

Chib, S. (1993). Bayes regression with autogressive errors: a Gibbs sampling approach. *Journal of Econometrics*, **58**, 275-294 .

Chib, S. and Greenberg E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, **85**, 347-361.

Chi, E. M. and Reinsel, G. C. (1989). Models for Longitudinal Data With Random Effects and AR(1) Errors. *Journal of the American Statistical Association*, **84**, 452-459.

Cox D. R. (1970). *The Analysis of Binary Data.* London, Chapman and Hall.

Cowles, M. K. (1996). Assessing and Accelerating Gibbs Sampler Convergence for Cumulative-Link Generalized Linear Models. *Statistics and Computing*, **6**, 101-111.

Crouchley, R. (1995). A random effects model for ordered categorical data. *Journal* *of the American Association* **90, 430,** 489-498.

Devroye, L. (1986). *Non-Uniform Random Variate Generation,* Springer-Verlag, New York.

Ezzet, F. and Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine.* **10**, 901-907.

Fienberg, S. E. (1994). *The Analysis of Cross-Classified Categorical Data Analysis.* MIT Press, Massachusetts.

Francom, S. F., Chuang, C and Landis, J. R. (1989). A log-linear model for ordinal data to characterize differential change among treatments. *Statistics in Medicine.* **8**, 571-582.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling Based Approach to Calculating Marginal Densities. *J. Am. Statist. Ass.*, **85**, 398-409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images . *IEEE Trans. Pattn. Anal. Math. Intell.*, **6**, 721-741.

Geyer, C. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.*, **7**, 437-482.

Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs Sampling. In *Bayesian Statistics* **4**, (eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. Smith), pp 642-649. Oxford: Oxford University Press.

Gilks, W. R.,Best, N. G. and Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs sampling. *Appl. Statist.* **44**, 455-472.

Gilks, W. R. and Wild P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Appl. Statist.* **41**, 337-348.

Gilmour, A. R., Anderson. R. D. and Rae A. L. (1985). The Analysis of Binomial Data by a Generalized Linear Mixed Model. *Biometrika* **72**, 593-599.

Grizzle, J. E. and Allen, D. M. (1969). Analysis of Growth and Dose Response Curves. *Biometrics*, **25**, 357-381.

Haber, M. (1985). Log-linear models for correlated marginal totals of a contingency table. *Communications in Statistics-Theory and Methods* **14** 2845-2856.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrics* **40**, 393-408.

Hedeker, D. and Gibbons, R. D. (1994). A Random Effects Ordinal Regression Model for Mutlilevel Analysis. *Biometrics* **50**, 933-944.

Hougaard, P. (1986a). Survival Models for Heterogenous Populations Derived from Stable Distributions. *Biometrika* **73** 387-396.

Jansen, J. (1990). On The Statistical Analysis of Ordinal Data when Extravariation is present. *Applied Statistics* **39**, 75-84.

Jones, B. and Kenward, M.G. (1989). *Design and Analysis of Cross-over Trials.* Chapman and Hall, London.

Kass R. E., Carlin B. P., Gelman, A. and Neal R. M. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician* **52, 2**, 93-100.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. Jr. and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* **33**, 133-158.

Kijewski, M.F., Swensson, R.G. and Judy, P.F. (1989). Analysis of Rating Data from Multiple -alternative Tasks. *Journal of Mathematical Psychology,* **33** 945-953.

Laird, N. M. and Ware, J. M. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.

Landis, J. R., Miller, M. E., Davis, C. S. and Koch, G. G. (1988). Some general methods for the analysis of categorical data in longitudinal studies. *Statistics in Medicine* **7**, 109-137.

Lewis S., Firth D. and Russell (1992). Carry-over effects in speech-transmission experiments. Progress Report, Dept. of Mathematics, University of Southampton.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13-22.

Liang, K. Y., Zeger, S. L. and Qaqish B. (1991). Multivariate Regression Analyses for Categorical Data. *J. R. Statist. Soc.* **B, 54, 1,** 3-40.

Lindstrom, M. and Bates, D. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.

McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449-453.

McCullagh, P. (1978). A Class of Parametric Models for the Analysis of Square Contingency Tables with Ordered Categories. *Biometrika*, **65**, 413-418.

McCullagh, P. (1980). Regression Models for Ordinal Data. *J. R. Statist. Soc.* **B, 42**, 109-142.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd Edition. Chapman and Hall, London.

Metropolis, N, Rosenbluth, A. W., Rosenbluth, M. N. Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine *J. Chem. Phys.,* **21**, 1087-1091.

Nandram, B. and Chen, M. H. (1996). Reparameterizing The Generalized Linear Model To Accelerate Gibbs Sampler Convergence. *J. Statist. Comput. Simul.*, **54**, 129-144.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *J. of the Royal. Statist. Society*, **Ser. A. 86**, 370-384.

Odell, P. L. and Feiveson, A. H. (1966). A Numerical Procedure to Generate a Sample Covariance Matrix. *Journal of the American Statistical Association*, **61**, 198-203.

Palmer, J. L. and Broemeling, L. D. (1993). Regression models with autocorrelation related errors: a Gibbs sampling approach. *American Statistical Association 1993 Proceedings of the Section on Bayesian Statistical Science*, 91-95.

Palmer, J. L. and Petit L. I. (1995). Risks of Using Improper Priors with Gibbs Sampling and Autocorrelated Errors. *Technical Report*, Dept. of Biomathematics, The University of Texas M. D., Houston.

Pan, J. X. and Thompson R. (1998). Quasi-Monte Carlo EM Algorithm for MLEs in Generalized Linear Mixed Models. In: *COMPSTAT98 Proceedings in Computational Statistics* (ed. R. Payne and P. Green), 419-424. Heidberg: Physica-Verlag.

Potthoff, R. F. and Roy, S. N. (1964). A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. *Biometrika*, **51**, 313-326.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.*

The Danish Institute of Educational Research. Copenhagen.

Rao, C. R. (1965). The Theory of Least Squares When the Parameters Are Stochastic and ITS Application to the Analysis of Growth Curve. *Biometrika*, **52**, 447-458.

Rao, C. R. (1967). Least Squares Theory Using an Estimated Dispersion Matrix and Its Application to Measurement of Signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1)*, Berkeley: University of California Press. 355-372.

Rubin, D. (1987). Comment on Paper by Tanner and Wong. *J. Am. Statist. Ass.*, **82**, 543-546.

Senn, S. (1993). . *Crossover Trials in Clinical Research*, Chichester: Wiley.

Smith, A.F.M. (1991). Bayesian Computational Methods. *Phil. Trans. R. Soc. Lond. A*. **337,** 369-386.

Stram, D. O., Wei, L. J. and Ware, J. H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time dependent covariates. *J. Am. Statist. Ass.* **83**, 3631-637.

Ten Have, T. and Uttal, D. (1994). Subject -specific and population-average continuation-ratio logit models for multiple discrete time survival profiles. *Applied Statistics* , **43**, 371-384.

Thompson, R. (1990). *Generalized linear models and applications to animal breeding. Advances in Statistical Methods for Genetic Improvement of livestock* , (ed. D. Gianola and K. Hammond) Berlin:Springer-Verlag, 312-327

Torgerson, W. S. (1958). *Theory and Methods of Scaling* , Wiley and Sons, New York.

Treviño-Villarreal M. L. (1999). Modelling the Information Content of Sovereign

Credit Ratings, *PhD Thesis*, Dept. of Management, University of Southampton.

Ware, J. H., Lipsitz, S. and Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**, 95-107.

Weddernburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear model. *Biometrics*, **63**, 27-32.

Williams, D. A. (1982). Extra-Binomial Variation in Logistic Linear Models. *Applied Statistics*, **31**, 144-148.

Williams, E. J. (1949). Experimental Designs balanced for the Estimation of Residual Effects of Treatment. *Australian Journal of Scientific Research*, **2**, 149-168.

Williams, E.J. (1950). Experimental Designs balanced for Pairs of Residual Effects. *Australian Journal of Scientific Research.*, **3** 351-363.

Wolfe R. (1996). Models and Estimation for Repeated Ordinal Responses, with Application to Telecommunication Experiments. *PhD Thesis*, Dept. of Mathematics, University of Southampton.

Zeger, S. L., Liang K. Y. and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.

Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with Random Effects: a Gibbs Sampling Approach. *J. Am. Statist. Ass*, **86**, 79-86.