Sampling Schemes for Generalized Linear Dirichlet Random Effects Models

Minjung Kyung* University of Florida

Jeff Gill† Washington University

George Casella‡ University of Florida

February 18, 2009

Abstract

We evaluate MCMC sampling schemes for a variety of link functions in generalized linear models with Dirichlet random effects. We find that models using Dirichlet process priors for the random effects tend to capture information in the data in a nonparametric fashion. In fitting the the Dirichlet process, dealing with the precision parameter has troubled model specifications in the past. Here we find that incorporating this parameter into the MCMC sampling scheme is not only computationally feasible, but also results in a more robust set of estimates, in that they are marginalized-over rather than conditioned-upon. Applications are provided with social science problems in areas where the data can be difficult to model. In all, we find that these models provide superior Bayesian posterior results in theory, simulation, and application.

*Postdoctoral Associate, Department of Statistics, University Florida, 102 Griffin-Floyd Hall, Gainesville, FL 32611. Supported by National Science Foundation Grants DMS-0631632 and SES-0631588.Email: kyung@stat.ufl.edu.

†Professor, Center for Applied Statistics, Washington University, One Brookings Dr., Seigle Hall, St. Louis, MO. Email: jgill@wustl.edu.

‡Distinguished Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Supported by National Science Foundation Grants DMS-04-05543, DMS-0631632 and SES-0631588. Email: casella@stat.ufl.edu.

1 Introduction

Generalized linear models (GLMs) have enjoyed considerable attention over the years, providing a flexible framework for modeling discrete responses using a variety of error structures. If we have observations that are discrete or categorical, $\mathbf{y} = (\mathbf{y_1}, \ldots, \mathbf{y_n})$, such data can often be assumed to be independent and from a distribution in the exponential family, where the likelihood function has components of the model such as the form of the link function and the type of error structures that result. The classic book by McCullagh and Nelder (1989) describes these models in detail; see also the more recent developments in Dey, Ghosh, and Mallick (2000) or Fahrmeir and Tutz (2001).

A generalized linear *mixed* model (GLMM) is an extension of a GLM that allows random effects, and can give us flexibility in developing a more suitable model when the observations are correlated, or where there may be other underlying phenomena that contribute to the resulting variability. Thus, the GLMM can be specified to accommodate outcome variables conditional on mixtures of possibly correlated random and fixed effects (Breslow and Clayton 1993, Buonaccorsi 1996, Wang, *et al.* 1998, Wolfinger and O'Connell, 1993). Details of such models, covering both statistical inferences and computational methods, can be found in the recent texts by McCulloch and Searle (2001) and Jiang (2007).

1.1 Sampling Schemes for GLMMs

There have been Markov chain Monte Carlo (MCMC) methods for the analysis of the GLMMs with random effects modeled with a normal distribution. Although the posteriors of parameters and the random effects are typically numerically intractable, especially when the dimension of the random effects is greater than one, there has been much progress in the development of sampling schemes. To handle the numerically intractable computation of the posteriors, Zeger and Karim (1991) proposed a Gibbs sampler, and considered rejection sampling with a Gaussian candidate for model parameters and the random effects. Improper priors were used, and it is possible that some choices of the hyperparameters could lead to an improper posterior distribution. Damien *et al.* (1999) proposed a Gibbs sampler using auxiliary variables for sampling non-conjugate and hierarchical models. Their methods are *slice sampling* methods derived from the full conditional posterior distribution. They mention that the assessment of convergence remains a major problem with the algorithm. However, Neal (2003) provided convergence properties of the posterior for slice sampling. Another sampling scheme was used by Chib *et al.* (1998) and Chib and Winkelmann (2001), who provided the Metropolis-Hastings (M-H) algorithms for various kinds of GLMMS. They proposed a multivariate-t distribution as a candidate density in an M-H implementation, taking the mean equal to the posterior mode, and variance equal to the inverse of the Hessian evaluated at the posterior mode.

1.2 Sampling Schemes for GLMDMs

Another variation of a GLMM was used by Dorazio, *et al.* (2007) and Gill and Casella (2008), where the random effects are modeled with a Dirichlet process, resulting in a generalized linear mixed Dirchlet model (GLMDM). Dorazio, *et al.* (2007) used a GLMDM with a log link for spatial heterogeneity in animal abundance. They proposed an empirical Bayesian approach with the Dirichlet process, instead of the regular assumption of normally distributed random effects, because they argued that for some species, the sources of heterogeneity in abundance is poorly understood or unobservable. They noted that the Dirichlet process prior is robust to errors in model specification and allows spatial heterogeneity in abundance to be specified in a data-adaptive way.

Gill and Casella (2008) suggested a GLMDM with an ordered probit link to model political science data, specifically modeling the stress, from public service, of Senate-confirmed political appointees as a reason for their short tenure. For the analysis, a semi-parametric Bayesian approach was adopted, using the Dirichlet process for the random effect. They noted that the use of a Dirichlet process mixture represents a new and useful paradigm for semi-informed prior information that reflects both information from observations and researcher intuition, where neither dominates.

Dirichlet process mixture models were introduced by Ferguson (1973) and Antoniak (1974). Blackwell and MacQueen (1973) showed that the marginal distribution of the Dirichlet process is equal to the distribution of the n^{th} step of a Polya urn process, and Korwar and Hollander (1973) characterized the joint distribution and looked at nonparametric empirical Bayes estimation of the distribution function based on Dirichlet process priors. Sethuraman (1994) showed that the Dirichlet measure is a distribution on the space of all probability measures, giving probability one to the subset of discrete probability measures. Turning to estimation, Lo (1984) derived the analytic form of a Bayesian density estimation that is generated by convoluting a known density kernel with a Dirichlet process, and Liu (1996) derived an identity for the profile likelihood estimator of m. However, Kyung, *et al.* (2008) looked at the properties of the MLE of m, and found that the likelihood function can be ill-behaved. They noted that incorporating a gamma prior, and using posterior mode estimation, results in a more stable solution. McAuliffe, Blei and Jordan (2006) used a similar strategy, using a posterior mean for the estimation of m.

Models with Dirichlet process priors are treated as hierarchical models in a Bayesian framework, and the implementation of these models through Bayesian computation and efficient algorithms has had much attention. Escobar and West (1995) provided a Gibbs sampling algorithm for the estimation of posterior distribution for all model parameters and the direct evaluation of predictive distributions, and also discussed inference about the precision parameter m using a gamma prior. MacEachern and Muller (1998) presented a Gibbs sampler with non-conjugate priors by using auxiliary parameters, and Neal (2000) provided an extended and more efficient Gibbs sampler to handle general Dirichlet process mixture models with non-conjugate priors by using a set of auxiliary parameters. Teh *et al.* (2006) also extended the auxiliary variable method of Escobar and West (1995) for posterior sampling of the precision parameter with a gamma prior. They developed hierarchical Dirichlet processes, with a Dirichlet prior for the base measure.

Kyung, *et al.* (2008) developed algorithms for estimation of the precision parameter and new MCMC algorithms for a linear mixed Dirichlet random effects models. Also, they showed how to extend the results to a generalized Dirichlet process mixed model with a probit link function. They derived a Gibbs sampler for the model parameters and the subclusters of the Dirichlet process, and used new parameterization of the hierarchical model to derive a Gibbs sampler that more fully exploits the structure of the model and mixes very well. They were also able to establish that the proposed sampler is an improvement, in terms of operator norm and efficiency, over other commonly used algorithms.

1.3 Summary

In this paper we develop MCMC sampling schemes for the generalized Dirichlet process mixture models by considering a GLMM with a general link function. For these models, we examine a Gibbs sampling method using auxiliary parameters, based on Damien *et al.* (1999), and a Metropolis-Hastings sampler with a Gaussian candidate from a log-linear model. We incorporate the precision parameter m into the Gibbs sampler, through the use of a gamma candidate distribution using a Laplace approximation for the calculation of the mean and variance of m, and use that in the gamma candidate. Comparatively, we find that the alternative slice sampler typically has higher autocorrelation in logistic regression and loglinear models than the proposed M-H algorithm.

Using the GLMDM with a general link function, Section 2 considers the generalized Dirichlet mixture model with Dirichlet process prior for the random effects instead of a normal assumption. In Section 3, we then estimate model parameters as well as the Dirichlet parameters using MCMC algorithms. This is used in Section 5 for simulation studies and data analysis. Section 7 summarizes these contributions and adds some perspective.

2 A Generalized Linear Mixed Dirichlet Model

Let $\mathbf{X}_i$ be covariates associated with the i^{th} observation, $\boldsymbol{\beta}$ be the coefficient vector, and ψ_i be a random effect accounting for subject-specific deviation from the underlying model. Assume that the $Y_i|\boldsymbol{\psi}$ are conditionally independent, each with a density from the exponential family, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$. Then, based on the notation on McCulloch and Searle (2001), the Generalized Linear Mixed Dirichlet Model (GLMDM) can be expressed as follows. First,

$$\begin{aligned} Y_i|\boldsymbol{\psi} &\overset{\text{ind}}{\sim} f_{Y_i|\boldsymbol{\psi}}(y_i|\boldsymbol{\psi}), \quad i = 1, \ldots, n \\ f_{Y_i|\boldsymbol{\psi}}(y_i|\boldsymbol{\psi}) &= \exp\left[\{y_i\gamma_i - b(\gamma_i)\}/\xi^2 - c(y_i, \xi)\right]. \end{aligned} \tag{1}$$

where y_i is discrete valued. Here, we know that $\mathrm{E}[Y_i|\boldsymbol{\psi}] = \mu_i = \partial b(\gamma_i)/\partial\gamma_i$. Moreover, using a link function $g(\cdot)$, we can express the transformed mean of Y_i, $\mathrm{E}[Y_i|\boldsymbol{\psi}]$, as a linear function,

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta} + \psi_i. \tag{2}$$

Here, for the Dirichlet process mixture models, we assume that

$$\begin{aligned} \psi_i &\sim G \\ G &\sim \mathcal{DP}(mG_0), \end{aligned} \tag{3}$$

where $\mathcal{DP}$ is the Dirichlet Process with base measure G_0 and precision parameter m. Blackwell and MacQueen (1973) proved that for $\psi_1, \ldots, \psi_n$ iid from $G \sim \mathcal{DP}$, the joint distribution of $\boldsymbol{\psi}$ is a product of successive conditional distributions of the form:

$$\psi_i|\psi_1, \ldots, \psi_{i-1}, m \sim \frac{m}{i-1+m}\, g_0(\psi_i) + \frac{1}{i-1+m}\sum_{l=1}^{i-1}\delta(\psi_l = \psi_i) \tag{4}$$

where $\delta()$ denotes the Dirac delta function and $g_0()$ is the density function of base measure.

We define a *subcluster* C to be a partition of the sample of size n into k groups, $k = 1, \ldots, n$, and we call these subclusters since the grouping is done nonparametrically rather than on substantive criteria. That is, the subclustering assigns different normal parameters across groups and the same parameters within groups: cases are iid only if they are assigned to the same subcluster.

Applying Lo (1984) Lemma 2 and Liu (1996) Theorem 1 to equation (4), we can calculate the likelihood function, which by definition is integrated over the random effects, as

$$L(\theta \mid \mathbf{y}) = \frac{\Gamma(m)}{\Gamma(m+n)}\sum_{k=1}^{n} m^k \sum_{C:|C|=k}\prod_{j=1}^{k}\Gamma(n_j)\int f(\mathbf{y}_{(j)}\,|\theta,\ \psi_j)\mathrm{d}dG_0(\psi_j),$$

where C defines the subclusters, $\mathbf{y}_{(j)}$ is the vector of y_is that are in subcluster j, and ψ_j is the common parameter for that subcluster. There are $\mathcal{S}_{n,k}$ different subclusters C, the Stirling Number of the Second Kind (Abramowitz and Stegun 1972, 824-825).

Here, we consider an $n \times k$ matrix A defined by

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

where a_i is a $1 \times k$ vector of all zeros except for a 1 in one position that indicates which group the observation is from. In other words, each column of matrix A represents a partition of the sample of size n into k groups. If the subcluster C is partitioned into groups $\{S_1, \ldots, S_k\}$, then if $i \in S_j$, $\psi_i = \eta_j$ and the random effect can be rewritten as

$$\boldsymbol{\psi} = A\boldsymbol{\eta}, \tag{5}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$ and $\eta_j \overset{iid}{\sim} G_0$ for $j = 1, \ldots, k$. Details of the Dirichlet process mixture models with the indicator matrix A are discussed in Kyung *et al.* (2008).

In this paper, we consider the models for the binary responses with probit and logit link function, and for count data with a log link function. First, for the binary responses,

$$Y_i \sim \text{ Bernoulli}(p_i), \quad i = 1, \ldots, n$$

where y_i is 1 or 0, thus $p_i = \mathrm{E(Y_i)}$ is the probability of a success for the i^{th} observation. Using a general link function (2) leads to a sampling distribution

$$f(\mathbf{y}|A) = \int \prod_{i=1}^{n} \left[g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i} \left[1 - g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{1-y_i} \mathrm{d}G_0(\boldsymbol{\eta}),$$

which typically can only be evaluated numerically. Examples of general link functions for binary outcomes are

$$\begin{array}{rlll} p_i = & g_1^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) = & \Phi(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) & \text{Probit} \\ p_i = & g_2^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) = & (1 + \exp(-\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{A}\boldsymbol{\eta})_i))^{-1} & \text{Logistic} \\ p_i = & g_3^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) = & 1 - \exp\left(-\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right) & \text{Cloglog} \end{array}$$

where $\Phi()$ is the cumulative distribution function of standard normal distribution.

For counting process data,

$$Y_i \sim \text{ Poisson}(\lambda_i), \quad i = 1, \ldots, n$$

where y_i is $0, 1, \ldots$, $\lambda_i = \mathrm{E(Y_i)}$ is the expected number of events for the i^{th} observation. Here, using a log link function

$$\lambda_i = g_3^{-1}(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i) = \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i),$$

a sampling distribution is

$$f(\mathbf{y}|A) = \prod_{i=1}^{n} \frac{1}{y_i!} \int \prod_{i=1}^{n} \exp\left\{-\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right\} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i} G_0(\boldsymbol{\eta})\, d\boldsymbol{\eta}.$$

3 Sampling Schemes for the Model Parameters

An overview of the general sampling scheme is as follows. We have three groups of parameters:

(i) m, the precision parameter of the Dirichlet process,

(ii) $\mathbf{A}$, the indicator matrix of the partition defining the subclusters, and

(iii) $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$, the model parameters.

We iterate between these three groups until convergence:

1. Conditional on m and A, generate $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)|\mathbf{A}, m$;
2. Conditional on $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$ and m, generate A, a new subcluster matrix;
3. Conditional on $(\boldsymbol{\eta}, \boldsymbol{\beta}, \tau^2)$ and A, generate m, the new precision parameter.

For the model parameters we add the priors

$$\begin{aligned} \beta|\sigma^2 &\sim N(\mathbf{0}, d^*\sigma^2 I) \\ \tau^2 &\sim \text{Inverse Gamma}(a, b), \end{aligned} \tag{6}$$

where $d^* > 1$ and (a, b) are fixed such that the inverse gamma is diffuse ($a = 1$, b very small). We can either fix σ^2 or put a prior on it and estimate it in the hierarchical model with priors, and here we will a value for σ^2. For the base measure of the Dirichlet process, we assume a normal distribution with mean 0 and variance τ^2, $N(0, \tau^2)$.

In the following sections we consider a number of sampling schemes go the estimation of the model parameters of a GLMDM. We will then turn to generation of the subclusters and the precision parameter.

3.1 Probit Models

Albert and Chib (1993) showed how truncated Normal sampling could be used to implement the Gibbs sampler for a probit model for binary responses. They use a latent variable U_i such that

$$U_i = X_i\beta + \psi_i + \eta_i, \quad \eta_i \sim N(0, \sigma^2), \tag{7}$$

and

$$y_i = 1 \quad \text{if } U_i > 0 \text{ and } y_i = 0 \quad \text{if } U_i \leq 0$$

for $i = 1, \ldots, n$. It can be shown that Y_i are independent Bernoulli random variables with the probability of success, $p_i = \Phi((X_i\beta - \psi_i)/\sigma)$, and without loss of generality, we fix $\sigma = 1$.

For given A and $\mathbf{U} = (U_1, \ldots, U_n)$, the likelihood function of model parameters and the latent variable is given by

$$\begin{aligned} L_k(\beta, \tau^2, \eta, \mathbf{U}|A, \mathbf{y}, \sigma^2) &= \prod_{i=1}^{n} \{I(U_i > 0)I(y_i = 1) + I(U_i \leq 0)I(y_i = 0)\} \\ &\times \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}|\mathbf{U} - X\beta - A\eta|^2} \left(\frac{1}{2\pi\tau^2}\right)^{k/2} e^{-\frac{1}{2\tau^2}|\eta|^2}. \end{aligned}$$

Let m and A be considered fixed for the moment. With priors given in (6) , the joint posterior distribution of $(\beta, \tau^2, \eta, \mathbf{U})$ given the data $\mathbf{y}$ is

$$\begin{aligned} \pi_k(\beta, \tau^2, \eta, \mathbf{U}|A, \mathbf{y}, \sigma^2) \propto \prod_{i=1}^{n} &\{I(U_i > 0)I(y_i = 1) + I(U_i \leq 0)I(y_i = 0)\} \\ &\times e^{-\frac{1}{2\sigma^2}|\mathbf{U} - X\beta - A\eta|^2} \left(\frac{1}{\tau^2}\right)^{k/2} e^{-\frac{1}{2\tau^2}|\eta|^2} e^{-\frac{1}{2d^*\sigma^2}|\beta|^2} \left(\frac{1}{\tau^2}\right)^{a+1} e^{-\frac{b}{\tau^2}}. \end{aligned}$$

Then for fixed m and A, it is straightforward to implement a Gibbs sampler using the full conditionals. Details are discussed in Appendix A.1

3.2 Logistic Models

We look at two samplers for the logistic model. The first is based on the slice sampler of Damien *et al.* (1999), while the second exploits a mixture representation of the logistic distribution; see Andrews and Mallows (1974) or West (1987).

3.2.1 Slice Sampling

The idea behind the slice sampler is the following. Suppose that the density $f(\theta) \propto l(\theta)\pi(\theta)$, where $l(\theta)$ is the likelihood and $\pi(\theta)$ is the prior, and it is not possible to sample directly from

$f(\theta)$. Using a latent variable U, define the joint density of θ and U by

$$f(\theta, u) \propto I\left\{u < l(\theta)\right\} \pi(\theta).$$

Then, $U|\theta$ is uniform $\mathcal{U}\left\{0, l(\theta)\right\}$, and $\theta|U = u$ is π restricted to the set $A_u = \{\theta : l(\theta) > u\}$.

The likelihood function of binary responses with logit link function can be written as

$$\begin{aligned} L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}|A, \mathbf{y}) &= \prod_{i=1}^{n} \left[\frac{1}{1+\exp(-\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{A}\boldsymbol{\eta})_i)}\right]^{y_i} \left[\frac{1}{1+\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)}\right]^{1-y_i} \\ &\quad \times \prod_{j=1}^{k} \left(\frac{1}{2\pi\tau^2}\right)^{1/2} \exp\left(-\frac{1}{2\tau^2}\eta_j^2\right), \end{aligned} \tag{8}$$

and if we introduce latent variables $\mathbf{U} = (U_1, \ldots, U_n)$ and $\mathbf{V} = (V_1, \ldots, V_n)$, we have the likelihood of the model parameters and the latent variables to be

$$\begin{aligned} &L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}|A, \mathbf{y}) \\ &= \prod_{i=1}^{n} I\left[u_i < \left\{\frac{1}{1+\exp(-\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{A}\boldsymbol{\eta})_i)}\right\}^{y_i},\ v_i < \left\{\frac{1}{1+\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)}\right\}^{1-y_i}\right] \\ &\quad \times \prod_{j=1}^{k} \left(\frac{1}{2\pi\tau^2}\right)^{1/2} \exp\left(-\frac{1}{2\tau^2}\eta_j^2\right) \end{aligned} \tag{9}$$

Thus, with priors that are given above, the joint posterior distribution of $(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V})$ can be expressed as

$$\begin{aligned} \pi_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}|A, \mathbf{y}) &\propto L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}|A, \mathbf{y}) \\ &\quad \times \prod_{j=1}^{k} \exp\left(-\frac{\eta_j^2}{2\tau^2}\right) \left(\frac{1}{\tau^2}\right)^{\frac{k}{2}+a+1} \exp\left(-\frac{b}{\tau^2}\right) \exp\left(-\frac{|\boldsymbol{\beta}|^2}{2d^*\sigma^2}\right). \end{aligned} \tag{10}$$

Then for fixed m and A, we can implement a Gibbs sampler using the full conditionals. Details are discussed in Appendix A.2.

3.2.2 A Mixture Representation

Next we consider a Gibbs sampler by using truncated normal variables in a manner that is similar to the Gibbs sampler of the probit models, which arise from a mixture representation of the logistic distribution. Andrews and Mallows (1974) discussed necessary and sufficient conditions under which a random variable Y may be generated as the ratio Z/V where Z and V are independent and Z has a standard normal distribution, and establish that when $V/2$ has the asymptotic distribution of the Kolmogorov distance statistic, Y is logistic. West (1987) generalized this result to the the exponential power family of distributions, showing them to be a subset of the class of scale mixtures of normals. The corresponding mixing distribution are explicitly obtained, identifying a close relationship between the exponential power family and a further class of normal scale mixtures, the stable distributions.

Based on Andrews and Mallows (1974), and West (1987), the logistic distribution is a scale mixture of a normal distribution with Kolmogorov-Smirnov distribution. From Devroye (1986), the Kolmogorov-Smirnov (K-S) density function is given by

$$f_X(x) = 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2 x e^{-2\alpha^2 x^2} \quad x \geq 0, \tag{11}$$

and we define the joint distribution

$$f_{Y,X}(y,x) = (2\pi)^{-1}\exp\left\{-\frac{1}{2}\left(\frac{y}{2x}\right)^2\right\} f_X(x)\frac{1}{2x}. \tag{12}$$

From the identities in Andrews and Mallows (1974) (see also Theorem 10.2.1 in Balakrishnan 1992), the marginal distribution of Y is then given by

$$f_Y(y) = \int_0^{\infty} f_{Y,X}(y,x)\mathrm{d}x = \sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha\exp\left(-\alpha|y|\right) = \frac{e^{-y}}{(1+e^{-y})^2}, \tag{13}$$

the density function of logistic distribution with mean 0 and variance $\frac{\pi^2}{3}$. Therefore, $Y \sim L\left(0, \frac{\pi^2}{3}\right)$, where $L()$ is the logistic distribution.

Now, using the likelihood function of binary responses with logit link function (8), consider the latent variable W_i such that

$$W_i = X_i\beta + \psi_i + \eta_i, \quad \eta_i \sim L\left(0, \frac{\pi^2}{3}\sigma^2\right), \tag{14}$$

with $y_i = 1 \;$ if $W_i > 0$ and $y_i = 0 \;$ if $W_i \leq 0$, for $i = 1, \ldots, n$. It can be shown that Y_i are independent Bernoulli random variables with $p_i = [1 + \exp(-\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{A}\boldsymbol{\eta})_i)]^{-1}$ the probability of success , and without loss of generality we fix $\sigma = 1$.

For given A, the likelihood function of model parameters and the latent variable is given by

$$\begin{aligned} L_k(\beta, \tau^2, \eta, \mathbf{U}|A, \mathbf{y}, \sigma^2) \;\; = \;\; & \prod_{i=1}^{n}\{I(U_i > 0)I(y_i = 1) + I(U_i \leq 0)I(y_i = 0)\} \\ & \times \int_0^{\infty}\left(\frac{1}{2\pi\sigma^2(2\xi)^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2(2\xi)^2}|\mathbf{U}-X\beta-A\eta|^2} \\ & \times 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}\mathrm{d}\xi\left(\frac{1}{2\pi\tau^2}\right)^{k/2} e^{-\frac{1}{2\tau^2}|\eta|^2}, \end{aligned}$$

where $\mathbf{U} = (U_1, \ldots, U_n)$, U_i is the truncated normal variable which is described in (7).

Let m and A be considered fixed for the moment. Thus, with priors given in (6) , the joint posterior distribution of $(\beta, \tau^2, \eta, \mathbf{U})$ given the data $\mathbf{y}$ is

$$\pi_k^L \propto \int_0^\infty L_k(\beta, \tau^2, \eta, \mathbf{U}|A, \mathbf{y}, \sigma^2) e^{-\frac{1}{2d^*\sigma^2}|\beta|^2} \left(\frac{1}{\tau^2}\right)^{a+1} e^{-\frac{b}{\tau^2}} \mathrm{d}\xi.$$

This representation avoids the problem of generating samples from the truncated logistic distribution, which is not easy to implement. As we now have the logistic distribution expressed as a normal mixture with the K-S distribution, we now only need to generate samples from the truncated normal distribution and the K-S distribution, and we can get a Gibbs sampler for the model parameters, The details are left to Appendix A.2.2.

3.3 Log Linear Models

Similar to Section 3.2, we look at two samplers for the loglinear model. The first is again based on the slice sampler of Damien *et al.* (1999), while the second is an M-H algorithm based on using a Gaussian density from log-transformed data as a candidate.

3.3.1 Slice Sampling

The likelihood function of the counting process data with log link function can be written as

$$\begin{aligned} L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}|A, \mathbf{y}) &= \prod_{i=1}^n \frac{1}{y_i!} e^{-\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i} \\ &\times \prod_{j=1}^k \left(\frac{1}{2\pi\tau^2}\right)^{1/2} \exp\left(-\frac{1}{2\tau^2}\eta_j^2\right), \end{aligned} \tag{15}$$

and the joint posterior distribution of $(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta})$ can be obtained by appending the priors for τ^2 and $\boldsymbol{\beta}$. As in Section 3.2.1 we introduce latent variables $\mathbf{U} = (U_1, \ldots, U_n)$ and $\mathbf{V} = (V_1, \ldots, V_n)$, yielding a likelihood of the model parameters and the latent variables, $L_k(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}|A, \mathbf{y})$, similar to (9). Setting up the Gibbs sampler is now straightforward, with details in Appendix A.3.1.

3.3.2 Metropolis-Hastings

Starting with the likelihood and priors described at (15), for the candidate distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, we consider the model

$$\begin{aligned} \log(Y_i) &= \mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2). \end{aligned}$$

which is a linear mixed Dirichlet model (LMDM). Sampling these model parameters is straightforward (Kyung *et al.* 2008), and the M-H algorithm is then easy to set up. Details are in Appendix A.3.2.

4 Sampling Schemes for the Dirichlet Parameters

4.1 Generating the Subclusters

We use a Metropolis-Hastings algorithm with a candidate taken from a multinomial/Dirichlet. This produces a Gibbs sampler that converges faster than the popular "stickbreaking" algorithm (Kyung *et al.* 2008).

For $t = 1, \dots T$, at iteration t

1. Starting from $(\theta^{(t)}, A^{(t)})$,
$$\theta^{(t+1)} \sim \pi(\theta \mid A^{(t)}, \mathbf{y}),$$
where $\theta = (\boldsymbol{\beta}, \tau^2, \eta)$ and the updating methods are discussed above.

2. Given $\theta^{(t+1)}$,
$$\begin{aligned} \mathbf{q}^{(t+1)} &= (q_1^{(t+1)}, \dots q_n^{(t+1)}) \sim \text{Dirichlet}(n_1^{(t)} + \beta_1, \dots, n_k^{(t)} + \beta_k, \beta_{k+1}, \dots, \beta_n) \\ A^{(t+1)} &\sim P(A')f(\mathbf{y}|\theta^{(t+1)}, A') \binom{n}{n_1 \ \cdots \ n_{k'}} \prod_{j=1}^{k'} [q_j^{(t+1)}]^{n'_j} \end{aligned} \tag{16}$$

where $n_j > 0$, $n_1 + \cdots + n_{k'} = n$.

Based on the value of the q_j in (16) we generate a candidate A that is an $n \times n$ matrix where each row is a multinomial, and the effective dimension of the matrix, the size of the subclusters, k, are the non-zero column sums. Deleting the columns with column sum zero is a marginalization of the multinomial distribution. The probability of the candidate follows

$$\begin{aligned} P\left(A^{(t+1)}\right) &= \frac{\Gamma(\sum_{j=1}^n \beta_j)}{\prod_{j=1}^n \Gamma(\beta_j)} \frac{\prod_{j=k^{(t)}}^n \Gamma(\beta_j)}{\Gamma(\sum_{j=k^{(t)}}^n \beta_j)} \int \prod_{j=1}^{k^{(t)}-1} q_j^{n_j^{(t)}+\beta_j-1} \left(1 - \sum_{j=k^{(t)}}^{n-1} q_j\right)^{n_k^{(t)}+\sum_{j=k^{(t)}}^n \beta_j - 1} d\, q_j \\ &= \frac{\Gamma(\sum_{j=1}^n \beta_j)}{\prod_{j=1}^{k^{(t)}-1} \Gamma(\beta_j)\Gamma(\sum_{j=k^{(t)}}^n \beta_j)} \frac{\prod_{j=1}^{k^{(t)}-1} \Gamma(n_j^{(t)} + \beta_j)\Gamma(n_k^{(t)} + \sum_{j=k^{(t)}}^n \beta_j)}{\Gamma(n + \sum_{j=1}^n \beta_j)} \end{aligned}$$

and the Metropolis-Hastings step is then done.

4.2 Gibbs Sampling the Precision Parameter

To estimate the precision parameter of the Dirichlet process, m, we start with the profile likelihood,

$$L(m \mid \theta, A, \mathbf{y}) = \frac{\Gamma(m)}{\Gamma(m+n)} m^k \prod_{j=1}^{k} \Gamma(n_j) f(\mathbf{y}|\theta, A). \tag{17}$$

Rather than estimate m, the better strategy is to include m in the Gibbs sampler, as the the maximum likelihood estimate from (17) can be very unstable (Kyung, *et al.* 2008). Using the prior $g(m)$ we get the posterior density

$$\pi(m \mid \theta, A, \mathbf{y}) = \frac{\frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k}{\int_0^\infty \frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k \mathrm{d}m}, \tag{18}$$

where $\int \pi(m \mid \theta, A, \mathbf{y})\ dm < \infty$ as long as the exponent of m is positive. Note also how far removed m is from the data, as the posterior only depends on the number of groups k. We consider a gamma distribution as a prior, $g(m) = m^{a-1} e^{-m/b} / \Gamma(a) b^a$, and generate m using an M-H algorithm with another gamma density as a candidate.

We choose the gamma candidate by using a approximate mean and variance of $\pi(m)$ to set the parameters of the candidate. To get the approximate mean and variance, we will use the Laplace approximation of Tierney and Kadane (1986). Applying their results, we have

$$\frac{\int m^\nu \frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k \mathrm{d}m}{\int \frac{\Gamma(m)}{\Gamma(m+n)} g(m) m^k \mathrm{d}m} \approx \sqrt{\frac{\ell_\nu''(\hat{m}_\nu)}{\ell''(\hat{m})}} \exp\left\{\ell_\nu(\hat{m}_\nu) - \ell(\hat{m})\right\}, \tag{19}$$

where

$$\begin{aligned}
\ell &= \log \frac{\Gamma(m)}{\Gamma(m+n)} + \log \frac{m^{a-1} e^{-m/b}}{\Gamma(a) b^a} + k \log m \\
\ell_\nu &= \ell + \nu \log m \\
\ell' &= \frac{\partial}{\partial m} \ell = \frac{1}{bm} \left[b(k+a-1) - m - bm \sum_{i=1}^{n} \frac{1}{m+i-1} \right] \\
{\ell_\nu}' &= \frac{\partial}{\partial m} \ell_\nu = \frac{1}{bm} \left[b(\nu+k+a-1) - m - bm \sum_{i=1}^{n} \frac{1}{m+i-1} \right] \\
\ell''(\hat{m}) &= \left. \frac{\partial^2}{\partial m^2} \ell \right|_{m=\hat{m}} = \frac{1}{b\hat{m}} \left[-1 - b \sum_{i=1}^{n} \frac{1}{\hat{m}+i-1} + b\hat{m} \sum_{i=1}^{n} \frac{1}{(\hat{m}+i-1)^2} \right] \\
{\ell_\nu}''(\hat{m}_\nu) &= \left. \frac{\partial^2}{\partial m^2} \ell_\nu \right|_{m=\hat{m}_\nu} = \ell''(\hat{m}_\nu),
\end{aligned}$$

where we get a simplification because the second derivative is evaluated at the zero of the first derivative. We use these approximations as the first and second moments of the candidate gamma distribution. Note that if $\hat{m} \approx \hat{m}_\nu$, then a crude approximation, which should be enough for Metropolis, is

$$\mathrm{Em}^\nu \approx (\hat{\mathrm{m}})^\nu .$$

5 Simulation Study

We evaluate our sampler through a number of simulation studies. We need to generate outcomes from Bernoulli or Poisson distributions with random effects that follow the Dirichlet process. To do this we fix K, then we set the parameter m according to the relation

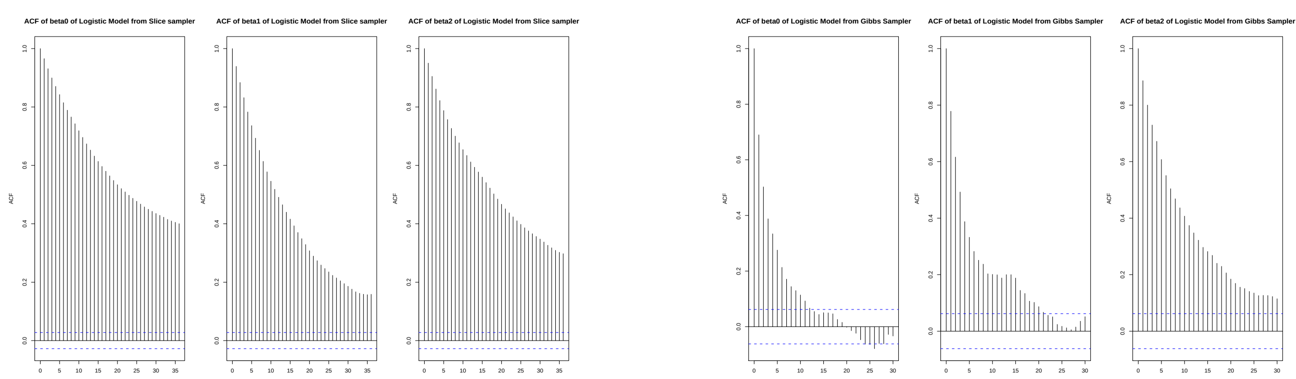
$$K = \sum_{i=1}^{n} \frac{m}{m+i-1}, \tag{20}$$

where we note that even if $\hat{m}$ is quite variable, there is less variability in $\hat{K} = \sum_{i=1}^{n} \frac{\hat{m}}{\hat{m}+i-1}$.

5.1 Logistic Models

Using the GLMDM with the logistic link function of Section 3.2, we set the parameters: $n = 100$, $K = 40$, $\tau^2 = 1$, and $\boldsymbol{\beta} = (1, 2, 3)$. Our Dirichlet process for the random effect has precision parameter m and base distribution $G_0 = N(0, \tau^2)$. Setting $K = 40$, yields $m = 24.21$. We then

Figure 1: ACF Plots of $\boldsymbol{\beta}$ for the GLMDM with logistic link. The left panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the slice sampler, and the right panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the K-S/normal mixture sampler.



Slice Sampler K-S Mixture Representation

Table 1: Estimation of the coefficients of the GLMDM with logistic link function and the estimate of K, with true values $K = 40$ and $\boldsymbol{\beta} = (1, 2, 3)$. Standard errors are in parentheses.

Estimation Method	β_0	β_1	β_2	K
Slice	1.4870(0.3139)	1.1512(0.3919)	2.6212(0.4632)	43. 4722(4.0477)
K-S Mixture	0.5444(0.1647)	0.7150(0.1874)	1.7462(0.2622)	43.1441(4.1668)

generated X_1 and X_2 independently from $N(0, 1)$, and used the fixed design matrix to generate the binary outcome Y. Then the Gibbs sampler was iterated 200 times to get values of m, A, $\boldsymbol{\beta}$, τ^2, η. This procedure was repeated 1000 times saving the last 500 draws as simulations from the posterior.

We compare the slice Gibbs sampler (**Slice**) to the Gibbs sampler with the K-S distribution normal scale mixture (**K-S Mixture**). For the estimation of K, we use the posterior mean of m, $\hat{m}$ and calculate $\hat{K}$ by using the equation (20). The starting points of $\boldsymbol{\beta}$ come from the maximum likelihood (ML) estimates using iteratively reweighted least squares: $(1.4623, 1.1263, 2.5547)$ and $(0.7705, 1.01509, 2.5134)$ for **Slice** and **K-S Mixture**, respectively. The numerical summary of this process is given in Table 1 and the ACF plots of $\boldsymbol{\beta}$ are given in Figure 2. The estimates of K were 43.4722 with standard error 4.0477 from **Slice** and 43.1441 with standard error 4.1668 from **K-S Mixture**. Obviously these turned out to be good estimates of the true $K = 40$. The estimate of $\boldsymbol{\beta}$ with **Slice** is closer to the true value than those with **K-S Mixture**. To evaluate the the convergence of $\boldsymbol{\beta}$, we consider the autocorrelation function (ACF) plots that are given in Figure 1. The Gibbs sampler of β from **Slice** is exhibits strong autocorrelation, implying poor mixing. Thus, if the Markov chain starts far from target distribution, we are not assured of reaching the stationary distribution in fixed time. Note, however, that the Gibbs sampler with normal scale mixture of K-S distribution mixes much better than that of **Slice**.

5.2 Log Linear Models

We now look at the GLMDM with the log link function of Section 3.3. The setting for the data generation is the same as the procedure that we discussed in the previous section except that we take $\boldsymbol{\beta} = (3, 0.5, 1)$. With $K = 40$, the solution of m from equation (20) is 24.21. As before, we generated X_1 and X_2 independently from $N(0, 1)$, and used the fixed design matrix to generate count data Y. The Gibbs sampler was iterated 200 times to produce draws of m, A, $\boldsymbol{\beta}$, τ^2, η. This procedure was repeated 1000 times, saving the last 500 values as draws from the posterior.

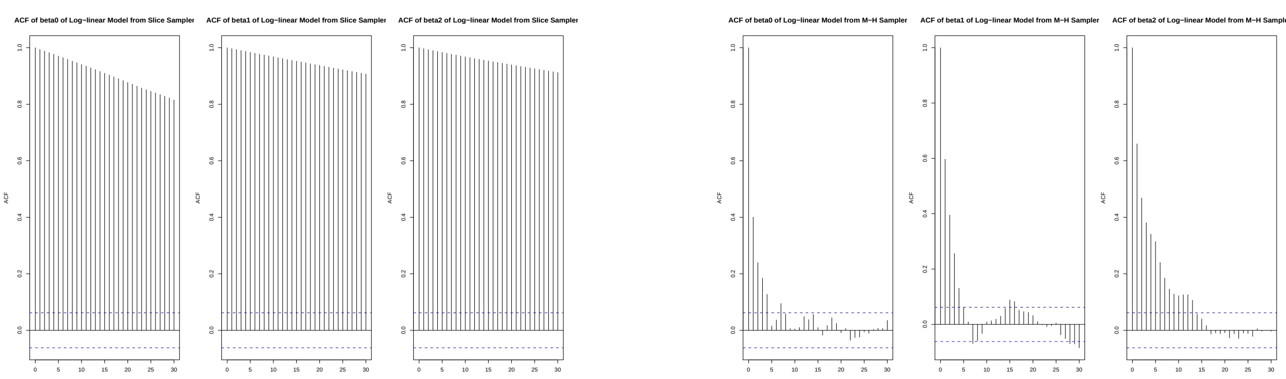
In this section, we compare the Gibbs sampler with the auxiliary variables (**Slice**) and the

Table 2: Estimation of the coefficients of the GLMDM with log link function and the estimate of K, with true values $K = 40$ and $\boldsymbol{\beta} = (3, 0.5, 1)$. Standard errors are in parentheses.

Estimation Method	β_0	β_1	β_2	K
Slice	2.5344(0.0048)	0.1249 (0.0573)	0.9141 (0.0105)	43.2432 (4.0129)
M-H Sampler	2.5848(0.1732)	0.3201(0.9680)	0.3685(0.9876)	43.4693(4.1617)

M-H sampler with a candidate density from the log-linear model (**M-H Sampler**). We use the posterior mean of m, $\hat{m}$, and calculate $\hat{K}$ by using (20) for the estimation of K. The starting points of $\boldsymbol{\beta}$ are set to the maximum likelihood (ML) estimates by using the iterative reweighted least squares. These were $(2.8321, -0.2541, 0.9402)$ and $(2.8675, -0.0058, -0.0788)$ for the Gibbs and the M-H sampler, respectively. The numerical summary is given in Table 2 and the ACF plots of $\boldsymbol{\beta}$ are given in Figure 2. The resulting estimates for K are 43.2432(4.0129) from **Slice** and 43.4693(4.1617) from **M-H Sampler**, which are fairly close to the true $K = 40$. The estimated β_1 from the **M-H Sampler** is near the true $\beta_1 = 0.5$, but the associated standard error is large. Conversely, the estimated β_2 from **Slice** is also near the true $\beta_2 = 1$ but has a smaller standard error. Once again, the consecutive draws of β of **Slice** from the Gibbs sampler are strongly autocorrelated. The convergence of $\boldsymbol{\beta}$ of **Slice** and **M-H Sampler** can be assessed by viewing the ACF plots in Figure 2. The M-H chain with candidate densities from log-linear models mixes better, giving additional confidence about convergence.

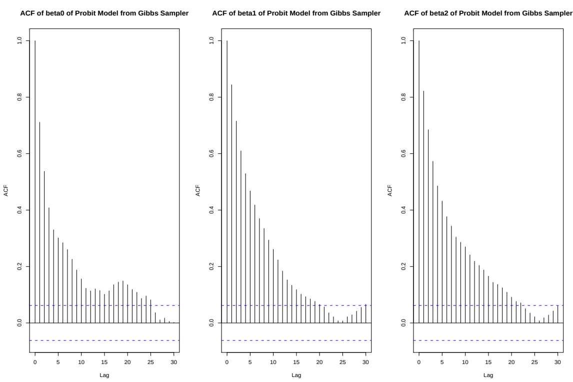
Figure 2: ACF Plots of $\boldsymbol{\beta}$ for the GLMDM with log link. The left panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the slice sampler, and the right panel are the plots for $(\beta_0, \beta_1, \beta_2)$ from the M-H sampler.



Gibbs Slice Sampler

Metropolis-Hastings Sampler

Figure 3: ACF Plots for $(\beta_0, \beta_1, \beta_2)$ for the GLMDM with probit link.



5.3 Probit Models

For completeness, we also generated data, similar to that described in Section 3.2, for a probit link. Here we only show the ACF plot from a latent variable Gibbs sampler as described in Section 3.1. where we see that the autocorrelations are not as good as the M-H algorithm, but better than those of the slice sampler.

6 Data Analysis

Save for the results in Figure 3, we have neglected empirical attention to the probit model. In this section we provide two real data examples that highlight the workings of generalized linear Dirichlet random effects models with a probit link function. Both examples are drawn from important questions in social science research: voting behavior and terrorism studies.

6.1 Social Attitudes in Scotland

The data for this example come from the Scottish Social Attitudes Survey, 2006 (UK Data Archive Study Number 5840). This study is based on face-to-face interviews conducted using computer assisted personal interviewing and a paper-based self-completion questionnaire, providing 1594 data points and 669 covariates. The focus was on attitudes towards government at the UK and national level, feelings about racial groups including discrimination, views on youth and youth crime, as well as exploring the Scottish sense of national identity.

Respondents were asked whether they favored full independence for Scotland with or without membership in the European Union versus remaining in the UK under different circumstances. This was used as a dichotomous outcome variable to explore the factors that contribute to advocating secession for Scotland. The explanatory variables used are: `househld` measuring the number of people living in the respondent's household, `rsex` for the respondent's sex (0=female, 1=male), `rage` giving the respondent's age, `relgsums` indicating identification with the Church of Scotland versus another or no religion, `ptyallgs` measuring party allegiance with the ordering of parties given from more conservative to more liberal, `idlosem` a dichotomous variable equal to one if the respondent agreed with the statement that increased numbers of Muslims in Scotland would erode the national identity, `marrmus`, another dichotomous variable equal to one if the respondent would be unhappy or very unhappy if a family member married a Muslim, `ukintnat` for agreement that the UK government works in Scotland's long-term interests, `voiceuk3` indicating that the respondent believes that the Scottish Parliament gives Scotland a greater voice in the UK, `nhssat` indicating satisfaction (1) or dissatisfaction (0) with the National Health Service, `hincdif2` a seven-point Likert scale showing the degree to which the respondent is living comfortably on current income or not (better in the positive direction), `unionsa` indicating union membership at work, `whrbrn` a dichotomous variable indicating birth in Scotland or not, and `hedqual2` the respondent's education level. We retain the variable names from the original study for ease of replication by others. All coding decisions are documented on a dedicated webpage.

We ran the Markov chain for 10,000 iterations saving the last 5,000 for analysis. All indications point towards convergence using empirical diagnostics (Geweke, Heidelberger & Welsh, graphics, etc.). The results in Table 3 are interesting in a surprising way. Notice that there are very similar results for the standard Bayesian probit model with flat priors (estimated in `JAGS`) and the GLMDM probit model, save for one coefficient (discussed below), which is attributable to the large sample size. This is a reassuring observation since it indicates that we are still operating in the general Bayesian construct.

Several of the coefficients point towards interesting findings from these results. Generally males, younger respondents, non-Church of Scotland adherents, union members, the better off financially, and so on, are more supportive of secession. The two variables regarding Muslims in Scotland turn out as expected: less tolerance of Muslims is positively associated with a desire to take Scotland out of the UK. The most interesting part of these results is the clear difference between the two modeling results for `whrbrn`. The regular probit model provides a statistically reliable result which indicates that native born Scots are *less* likely to support secession, which runs counter to research and generally accepted wisdom about nationalism and national identification. However, the GLMDM result has a statistically defensible result showing the *opposite* effect. Clearly there exists additional information in the data that are being captured in the second model. So with large datasets the results may not differ much, except for circumstances where the data have strong nonparametric contributions.

Table 3: Probit Models for Attitudes in Scotland

	Standard Probit				GLMDM Probit			
Coefficient	COEF	SE	95% CI		COEF	SE	95% CI	
`Intercept`	-0.369	0.225	-0.241	-0.497	-0.482	0.218	-0.908	-0.055
`househld`	-0.065	0.032	0.063	-0.193	-0.076	0.030	-0.135	-0.017
`rsex`	0.180	0.070	0.308	0.052	0.165	0.068	0.031	0.299
`rage`	-0.008	0.003	0.120	-0.135	-0.009	0.002	-0.014	-0.004
`relgsums`	-0.213	0.085	-0.085	-0.341	-0.210	0.085	-0.378	-0.042
`ptyallgs`	0.043	0.009	0.171	-0.085	0.042	0.009	0.025	0.059
`idlosem`	0.216	0.100	0.344	0.088	0.201	0.098	0.009	0.393
`marrmus`	0.210	0.088	0.338	0.082	0.204	0.086	0.035	0.373
`ukintnat`	-0.328	0.057	-0.200	-0.456	-0.324	0.057	-0.436	-0.213
`natinnat`	0.182	0.052	0.310	0.054	0.173	0.052	0.070	0.275
`voiceuk3`	0.096	0.039	0.224	-0.032	0.097	0.039	0.022	0.173
`nhssat`	0.197	0.074	0.325	0.069	0.183	0.074	0.039	0.327
`hincdif2`	-0.095	0.041	0.033	-0.223	-0.092	0.041	-0.172	-0.011
`unionsa`	0.181	0.069	0.309	0.053	0.238	0.080	0.080	0.395
`whrbrn`	-0.411	0.124	-0.283	-0.539	0.250	0.088	0.077	0.423
`hedqual2`	-0.059	0.018	0.068	-0.187	-0.064	0.018	-0.100	-0.028

6.2 Terrorism Targeting

In this example we look at terrorist activity in 22 Asian democracies over 8 years (1990-1997) with data subsetted from Koch and Cranmer (2007). Data problems (a recurrent problem in the empirical study of terrorism) reduce the number of cases to 162 and make fitting any standard model difficult due to the generally poor level of measurement. The outcome of interest is dichotomous, indicating whether not there was at least one violent terrorist act in a country/year pair. In order to control for the *level* of democracy (`DEM`) in these countries we use the Polity IV 21-point democracy scale ranging from -10 indicating a hereditary monarchy to +10 indicating a fully consolidated democracy (Gurr, Marshall, and Jaggers 2003). The variable `FED` is assigned zero if sub-national governments do not have substantial taxing, spending, and regulatory authority, and one otherwise. We look at three rough classes of government structure with the variable `SYS` coded as: (0) for direct presidential elections, (1) for strong president elected by assembly, and (2) dominant parliamentary government. Finally, `AUT` is a dichotomous variable

Table 4: Probit Models for Terrorism Incidents

	Standard Probit				GLMDM Probit			
Coefficient	COEF	SE	95% CI		COEF	SE	95% CI	
`Intercept`	0.249	0.337	0.035	0.463	0.127	0.188	-0.241	0.495
`DEM`	0.109	0.035	-0.105	0.323	0.058	0.019	0.020	0.095
`FED`	0.649	0.469	0.435	0.863	0.258	0.254	-0.241	0.756
`SYS`	-0.817	0.252	-1.031	-0.603	-0.420	0.137	-0.690	-0.151
`AUT`	1.619	0.871	1.406	1.833	0.450	0.371	-0.277	1.176

indicating whether or not there are autonomous regions not directly controlled by central government. The key substantive question evaluated here is whether specific structures of government and sub-governments lead to more or less terrorism.

We ran the Markov chain for 50,000 iterations disposing of the first half. There is no evidence of non-convergence in these runs using standard diagnostic tools. Table 4 again provides results from two approaches: a standard Bayesian probit model with flat priors, and a Dirichlet random effects model. Notice first that while there are no changes in sign or statistical reliability for the estimated coefficients, the magnitudes of the effects are uniformly smaller with the enhanced model: four of the estimates are roughly twice as large and the last one is about three times as large as in the standard model. This indicates that there is extra information in the data detected by the Dirichlet random effect that tends to dampen the size of the effect of these explanatory variables on explaining incidences of terrorist attacks. Specifically, running the standard probit model would find an *exaggerated* relationship between these explanatory variables and the outcome.

The results are also interesting substantively. The more democratic a country is, the more terrorist attacks they can expect. This is consistent with the literature in that autocratic nations tend to have more security resources per capita and fewer civil rights to worry about. Secondly, the more the legislature holds central power, the fewer expected terrorist attacks. This also makes sense, given what is known; disparate groups in society tend to have a greater voice in government when the legislature dominates the executive. Two results are puzzling and are therefore worth further investigation. Strong sub-governments and the presence of autonomous regions both lead to more expected terrorism.

7 Discussion

In this paper we demonstrate how to set up and run sampling schemes for the generalized linear mixed Dirichlet model with a variety of link functions. We focus on the mixed effects model with a Dirichlet process prior for the random effects instead of the normal assumption, as in standard approaches. We are able to estimate model parameters as well as the Dirichlet parameters using convenient MCMC algorithms, and to draw latent information from the data. Simulation studies and empirical studies demonstrate the effectiveness of this approach.

The major methodological contributions here are the derivation and evaluation of strategies of estimation for model parameters in Section 3 and the inclusion of the precision parameter directly into the Gibbs sampler for estimation in Section 4.2. In the latter case, including the precision parameter in the Gibbs sampler means that we are marginalizing over the parameter rather than conditioning on it leading to a more robust set of estimates. Moreover, we have seen a large amount of variability in the performance of MCMC algorithms, with the slice sampler typically being less optimal than either a K-S mixture representation or a Metropolis-Hastings algorithm.

Finally, we observed that the additional effort needed to include a Dirichlet process prior for the random effects in two empirical examples with social science data, which tends to be more messy and interrelated than that in other fields, added significant value to the data analysis. We found that the GLMDM model can find additional information in the data which affects parameter estimates. In particular, in the case of social attitudes in Scotland the GLMDM model corrected a clearly illogical finding in the usual probit analysis. For the second example, we found that the GLMDM specification dampened-down over enthusiastic findings from a conventional model. In both cases either non-Bayesian or Bayesian models with flat priors would have reported results that had substantively misleading findings.

8 References

Abramowitz, M. and Stegun, I. A. (1972). “Stirling Numbers of the Second Kind.” Section 24.1.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York: Dover, 824-825.

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

Andrews D. F. and Mallows, C. L. (1974). “Scale Mixtures of Normal Distributions.” *Journal of the Royal Statistical Society, Series B* **36**, 99-102.

Antoniak, Charles E. (1974). "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems." *Annals of Statistics* **2**, 1152-1174.

Balakrishnan, N. (1992). *Handbook of the Logistic Distribution.* CRC Press.

Blackwell, D. and MacQueen, J. B. (1973). "Discreteness of Ferguson Selections." *Annals of Statistics* **1**, 365-358.

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

Buonaccorsi, J. P. (1996). "Measurement Error in the Response in the General Linear Model." *Journal of the American Statistical Association* **91**, 633-642.

Chib, S., Greenberg, E., and Chen, Y. (1998). "MCMC methods for fitting and comparing multinomial response models." Technical Report, Economics Working Paper Archive, Washington University at St. Louis.

Chib, S. and Winkelmann, R. (2001). "Markov Chain Monte Carlo Analysis of Correlated Count Data." *Journal of Business and Economic Statistics* **19**, 428-435.

Damien, P., Wakefield, J., and Walker, S. (1999) "Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables." *Journal of the Royal Statistical Society, Series B* **61**, 331-344.

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

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective.* New York: Marcel Dekker.

Dorazio, R. M., Mukherjee, B., Zhang, L., Ghosh, M., Jelks, H. L. and Jordan, F. (2007). "Modelling Unobserved Sources of Heterogeneity in Animal Abundance Using a Dirichlet Process Prior." *Biometrics*, Online Publication August 3, 2007.

Escobar, M. D. and West, M. (1995). "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association* **90**, 577-588.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Second Edition. New York: Springer.

Ferguson, T. S. (1973). "A Bayesian analysis of Some Nonparametric Problems." *Annals of Statistics* **1**, 209-230.

Gill, J. and Casella, G. (2008). "Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation." Technical Report, University of Florida, Department of Statistics. To appear in *J. Amer. Statist. Assoc.*

Gurr, T. R., Marshall, M. G., and Jaggers, K. (2003). *PolityIV*, `http://www.cidcm.umd.edu/inscr/polity/`.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer-Verlag.

Koch, M. T. Cranmer, S. (2007). "Terrorism than Governments of the Right? Testing the 'Dick Cheney' Hypothesis: Do Governments of the Left Attract More than Goverments of the Right?" *Conflict Management and Peace Science* **24**, 311-326.

Korwar, R. M. and Hollander, M. (1973). "Contributions to the Theory of Dirichlet Processes." *Annals of Probability* **1**, 705-711.

Kyung, M, Gill, J,, and Casella G. (2008). "Dirichlet Random Effects Models." Technical Report, University of Florida, Department of Statistics. Available at http://www.stat.ufl.edu/∼casella/Papers/GLMDM3-Linear.pdf

Liu, J. S. (1996). "Nonparametric Hierarchical Bayes Via Sequential Imputations." *Annals of Statistics* **24**, 911-930.

Lo, A. Y. (1984). "On A Class of Bayesian Nonparametric Estimates: I. Density Estimates." *Annals of Statistics* **12**, 351-357.

MacEachern, S. N. and Müller, P. (1998). "Estimating Mixture of Dirichlet Process Model." *Journal of Computational and Graphical Statistics* **7**, 223-238.

McAuliffe, J. D., Blei, D. M. and Jordan, M. I. (2006). "Nonparametric Empirical Bayes for the Dirichlet Process Mixture Model." *Statistics and Computing* **16**, 5-14.

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

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* New York: John Wiley & Sons.

Neal, R. M. (2000). "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics* **9**, 249-265.

Neal, R. M. (2003). “Slice Sampling.” *The Annals of Statistics* **31**, 705-741.

Sethuraman, J. (1994). “A Constructive Definition of Dirichlet Priors.” *Statistica Sinica* **4**, 639-650.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006). “Hierarchical Dirichlet Processes.” *Journal of the American Statistical Association* **101**, 1566-1581.

Tierney, L. and Kadane, J. B. (1986). “Accurate Approximations for Posterior Moments and MArginal Densities.” *Journal of the American Statistical Association* **81**, 82-86.

Wang, N., X. Lin, R. G. Gutierrez, and R. J. Carroll. (1998). “Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models.” *Journal of the American Statistical Association* **93**, 249-261.

West, M. (1987). “On Scale Mixtures of Normal Distributions”, *Biometrika* **74**, 646-648.

Wolfinger, R. and M. O’Connell. (1993). “Generalized Linear Mixed Models: A Pseudolikelihood Approach.” *Journal of Statistical Computation and Simulation* **48**, 233-243.

Zeger, S. M. and R. Karim (1991). “Generalized Linear Models With Random Effects; A Gibbs Sampling Approach.” *Journal of the American Statistical Association* **86**, 79-86.

A Generating the Model Parameters

A.1 A Probit Model

The Gibbs sampler of $(\beta, \tau^2, \eta, \mathbf{U})$ is obtained by iterating through:

$$
\begin{aligned}
\eta|\beta, \tau^2, \mathbf{U}, A, \mathbf{y} &\sim N_k\left(\frac{1}{\sigma^2}\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1} A'(\mathbf{U} - X\beta), \left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1}\right) \\
\beta|\tau^2, \eta, \mathbf{U}, A, \mathbf{y} &\sim N_p\left(\left(\frac{1}{d^*}I + X'X\right)^{-1} X'(\mathbf{U} - A\eta), \sigma^2\left(\frac{1}{d^*}I + X'X\right)^{-1}\right) \\
\tau^2|\beta, \eta, \mathbf{U}, A, \mathbf{y} &\sim \text{Inverse Gamma}\left(\frac{k}{2} + a, \frac{1}{2}|\eta|^2 + b\right) \\
U_i|\beta, \tau^2, \eta, A, \mathbf{y} &\sim \begin{cases} N\left(X_i\beta + (A\eta)_i, \sigma^2\right) I(U_i > 0) & \text{if } y_i = 1 \\ N\left(X_i\beta + (A\eta)_i, \sigma^2\right) I(U_i \leq 0) & \text{if } y_i = 0 \end{cases}
\end{aligned}
$$

Note that we can marginalize out η yielding

$$\begin{aligned} L_k(\beta, \sigma^2, \tau^2, \mathbf{U}|A, \mathbf{y}) &= \int L_k(\beta, \sigma^2, \tau^2, \eta, \mathbf{U}|A, \mathbf{y}) \mathrm{d}\eta \\ &= \prod_{i=1}^{n} \{ I(U_i > 0) I(y_i = 1) + I(U_i \le 0) I(y_i = 0) \} \\ &\qquad \times \frac{|A^*|^{-1/2}}{(2\pi\sigma^2)^{n/2}(\tau^2)^{k/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{U}-X\beta)'\left[I - \frac{1}{\sigma^2}A(A^*)^{-1}A'\right](\mathbf{U}-X\beta)}, \end{aligned}$$

where $A^* = \frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A$. Note that $|A^*|^{-1/2} = |I - \frac{1}{\sigma^2}A(A^*)^{-1}A'|^{-1/2}$.

A.2 A Logistic Model

A.2.1 Slice Sampling

For fixed m and A, a Gibbs sampler of $(\boldsymbol{\beta}, \tau^2, \eta, \mathbf{U}, \mathbf{V})$ is

- for $d = 1, \ldots, p$,

$$\beta_d | \boldsymbol{\beta}_{-d}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim \begin{cases} N\left(0, d^*\sigma^2\right) & \text{if } \beta_d \in \left[\left\{\max\left(\max_{X_{id}>0}\left(\frac{\alpha_{id}}{X_{id}}\right)\right), \left(\max_{X_{id}\le 0}\left(\frac{\gamma_{id}}{X_{id}}\right)\right)\right\}, \right. \\ & \qquad \left. \left\{\min\left(\min_{X_{id}\le 0}\left(\frac{\alpha_{id}}{X_{id}}\right)\right), \left(\min_{X_{id}>0}\left(\frac{\gamma_{id}}{X_{id}}\right)\right)\right\}\right] \\ 0 & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} \alpha_{id} &= -\log\left(u_i^{-\frac{1}{y_i}} - 1\right) - \sum_{l \ne d} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i \quad \text{for } i \in S \\ \gamma_{id} &= \log\left(v_i^{\frac{1}{y_i - 1}} - 1\right) - \sum_{l \ne d} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i \quad \text{for } i \in F. \end{aligned}$$

Here, $S = \{i : y_i = 1\}$ and $F = \{i : y_i = 0\}$.

- $\tau^2 | \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim$ Inverse Gamma $\left(\frac{k}{2} + a, \frac{1}{2}|\boldsymbol{\eta}|^2 + b\right)$

- for $j = 1, \ldots, k$,

$$\eta_j | \boldsymbol{\beta}, \tau^2, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim \begin{cases} N\left(0, \tau^2\right) & \text{if } \eta_j \in \left(\max_{i \in S_j}\{\alpha_i^*\}, \min_{i \in S_j}\{\gamma_i^*\}\right) \\ 0 & \text{otherwise} \end{cases},$$

where

$$\begin{aligned} \alpha_i^* &= -\log\left(u_i^{-1} - 1\right) - \mathbf{X}_i\boldsymbol{\beta} \quad \text{for } i \in S \\ \gamma_i^* &= \log\left(v_i^{-1} - 1\right) - \mathbf{X}_i\boldsymbol{\beta} \quad \text{for } i \in F \end{aligned}$$

- for $i = 1, \ldots, n$,

$$\begin{aligned}
\pi_k(U_i|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{V}, A, \mathbf{y}) &\propto I\left[u_i < \left\{\frac{1}{1+\exp(-\mathbf{X}_i\boldsymbol{\beta} - \eta_j)}\right\}^{y_i}\right] \quad \text{for } i \in S \\
\pi_k(V_i|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, A, \mathbf{y}) &\propto I\left[v_i < \left\{\frac{1}{1+\exp(\mathbf{X}_i\boldsymbol{\beta} + \eta_j)}\right\}^{1-y_i}\right] \quad \text{for } i \in F.
\end{aligned}$$

A.2.2 K-S Mixture

Given ξ, for fixed m and A, a Gibbs sampler of $(\beta, \tau^2, \eta, \mathbf{U})$ is

$$\begin{aligned}
\eta|\beta, \tau^2, \mathbf{U}, A, \mathbf{y}, \sigma^2 &\sim N_k\left(\frac{1}{\sigma^2(2\xi)^2}\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2(2\xi)^2}A'A\right)^{-1}A'(\mathbf{U} - X\beta),\right. \\
&\qquad \left.\times\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2(2\xi)^2}A'A\right)^{-1}\right) \\
\beta|\tau^2, \eta, \mathbf{U}, A, \mathbf{y}, \sigma^2 &\sim N_p\left(\left(\frac{1}{d^*}I + \frac{1}{(2\xi)^2}X'X\right)^{-1}\frac{1}{(2\xi)^2}X'(\mathbf{U} - A\eta),\right. \\
&\qquad \left.\times\sigma^2\left(\frac{1}{d^*}I + \frac{1}{(2\xi)^2}X'X\right)^{-1}\right) \\
\tau^2|\beta, \eta, \mathbf{U}, A, \mathbf{y}, \sigma^2 &\sim \text{Inverse Gamma}\left(\frac{k}{2} + a, \frac{1}{2}|\eta|^2 + b\right) \\
U_i|\beta, \tau^2, \eta, A, y_i, \sigma^2 &\sim \begin{cases} N\left(X_i\beta + (A\eta)_i, \sigma^2(2\xi)^2\right) I(U_i > 0) & \text{if } y_i = 1 \\ N\left(X_i\beta + (A\eta)_i, \sigma^2(2\xi)^2\right) I(U_i \leq 0) & \text{if } y_i = 0 \end{cases}
\end{aligned}$$

Then we update ξ from

$$\xi|\beta, \tau^2, \eta, \mathbf{U}, A, \mathbf{y} \sim \left(\frac{1}{(2\xi)^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2(2\xi)^2}|\mathbf{U} - X\beta - A\eta|^2} 8\sum_{\alpha=1}^{\infty}(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}.$$

The conditional posterior density of ξ is the product of a inverse Gamma with parameters $\frac{\alpha}{2} - 1$ and $-\frac{1}{8\sigma^2}|\mathbf{U} - X\beta - A\eta|^2$, and the infinite sum of the sequence $(-1)^{\alpha+1}\alpha^2\xi e^{-2\alpha^2\xi^2}$. To generate samples from this target density, we consider the alternating series method that is proposed by Devroye (1986). Based on his notation, we take

$$\begin{aligned}
ch(\xi) &= 8\left(\frac{1}{\xi^2}\right)^{n/2} e^{-\frac{1}{8\sigma^2\xi^2}|\mathbf{U} - X\beta - A\eta|^2}\xi e^{-2\xi^2} \\
a_n(\xi) &= (\alpha + 1)^2 e^{-2\xi^2\{(\alpha+1)^2 - 1\}}
\end{aligned}$$

Here, we need to generate sample from $h(\xi)$, and we use accept-reject sampling with candidate $g(\xi^*) = 2e^{-2\xi^*}$, the exponential distribution with $\lambda = 2$, where $\xi^* = \xi^2$. Then we follow Devroye's method.

A.3 A Log Link Model

A.3.1 Slice Sampling

Starting from the likelihood $L(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V})$, and the priors on $(\boldsymbol{\beta}, \tau^2)$, we have the following Gibbs sampler of the model parameters.

- The conditional posterior distribution of $\boldsymbol{\beta}$:

$$\pi_K(\boldsymbol{\beta}|\tau^2, \boldsymbol{\eta}, A, \mathbf{y}, \mathbf{U}, \mathbf{V}) \propto e^{-\frac{1}{2d^*\sigma^2}|\boldsymbol{\beta}|^2} \times \prod_{i=1}^{n} I\left[u_i < \exp\left\{y_i(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right\}, v_i > \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right].$$

For $d = 1, \ldots, p$,

$$\pi_K(\beta_d|\boldsymbol{\beta}_{-d}\tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto e^{-\frac{1}{2d^*\sigma^2}\beta_d^2} \times \prod_{i=1}^{n} I\left[u_i < \exp\left\{y_i(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right\}, v_i > \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right],$$

which can be expressed as:

$$\pi_K(\beta_d|\boldsymbol{\beta}_{-d}\tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto e^{-\frac{1}{2d^*\sigma^2}\beta_d^2} \times \prod_{i=1}^{n} I\left[X_{id}\beta_d < \frac{1}{y_i}\log(u_i) - \sum_{l\neq j} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i, X_{id}\beta_d < \log(v_i) - \sum_{l\neq j} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i\right],$$

where $\boldsymbol{\beta}_{-d} = (\beta_1, \ldots, \beta_{d-1}, \beta_{d+1}, \ldots, \beta_p)$. The full conditional posterior of β_d for $d = 1, \ldots, p$ is

$$\pi_k(\beta_d|\boldsymbol{\beta}_{-j}\tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto e^{-\frac{1}{2d^*\sigma^2}\beta_d^2}\beta_d \in \left[\left\{\max\left(\max_{X_{id}>0}\left(\frac{\alpha_{id}^*}{X_{id}}\right)\right), \left(\max_{X_{id}\leq 0}\left(\frac{\gamma_{id}^*}{X_{id}}\right)\right)\right\}, \left\{\min\left(\min_{X_{id}\leq 0}\left(\frac{\alpha_{id}^*}{X_{id}}\right)\right), \left(\min_{X_{id}>0}\left(\frac{\gamma_{id}^*}{X_{id}}\right)\right)\right\}\right],$$

where

$$\begin{aligned} \alpha_{id}^* &= \frac{1}{y_i}\log(u_i) - \sum_{l\neq d} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i \quad \text{for } i \in S \\ \gamma_{id}^* &= \log(v_i) - \sum_{l\neq d} X_{il}\beta_l - (\mathbf{A}\boldsymbol{\eta})_i \quad \text{for } i \in F \end{aligned}$$

Thus, for $d = 1, \ldots, p$,

$$\beta_d|\boldsymbol{\beta}_{-d}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim$$
$$\begin{cases} N\left(0, d^*\sigma^2\right) & \text{if } \ \beta_d \in \begin{array}{l} \left[\left\{\max\left(\max_{X_{id}>0}\left(\frac{\alpha^*_{id}}{X_{id}}\right)\right), \left(\max_{X_{id}\leq 0}\left(\frac{\gamma^*_{id}}{X_{id}}\right)\right)\right\},\right. \\ \left.\left\{\min\left(\min_{X_{id}\leq 0}\left(\frac{\alpha^*_{id}}{X_{id}}\right)\right), \left(\min_{X_{id}>0}\left(\frac{\gamma^*_{id}}{X_{id}}\right)\right)\right\}\right] \end{array} \\ 0 & \text{otherwise} \end{cases}$$

- The conditional posterior distribution of τ^2:

$$\pi_k(\tau^2|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto \left(\frac{1}{\tau^2}\right)^{k/2+a+1} e^{-\frac{1}{\tau^2}\left(\frac{1}{2}|\boldsymbol{\eta}|^2+b\right)}.$$

Thus,

$$\tau^2|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim \text{Inverse Gamma}\left(\frac{k}{2} + a, \frac{1}{2}|\boldsymbol{\eta}|^2 + b\right).$$

- The conditional posterior distribution of $\boldsymbol{\eta}$:

$$\pi_k(\boldsymbol{\eta}|\boldsymbol{\beta}, \tau^2, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto \prod_{j=1}^{k} e^{-\frac{1}{2\tau^2}\eta_j^2} \prod_{i\in S_j} I\left[u_i < \exp\left\{y_i(\mathbf{X}_i\boldsymbol{\beta} + \eta_j)\right\}, v_i > \exp(\mathbf{X}_i\boldsymbol{\beta} + \eta_j)\right].$$

For $j = 1, \ldots, k$,

$$\begin{aligned} \pi_k(\eta_j|\boldsymbol{\beta}, \tau^2, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) &\propto e^{-\frac{1}{2\tau^2}\eta_j^2} \prod_{i\in S_k} I\left[u_i < \exp\left\{y_i(\mathbf{X}_i\boldsymbol{\beta} + \eta_j)\right\}, v_i > \exp(\mathbf{X}_i\boldsymbol{\beta} + \eta_j)\right] \\ &\propto e^{-\frac{1}{2\tau^2}\eta_j^2} I\left[\eta_j \in \left(\max_{i\in S_k}\left\{\gamma_i^*\right\}, \min_{i\in S_k}\left\{\alpha_i^*\right\}\right)\right], \end{aligned}$$

where

$$\begin{aligned} \alpha_i^* &= \frac{1}{y_i}\log(u_i) - \mathbf{X}_i\boldsymbol{\beta} \\ \gamma_i^* &= \log(v_i) - \mathbf{X}_i\boldsymbol{\beta} \end{aligned}$$

Thus, for $j = 1, \ldots, k$,

$$\eta_j|\boldsymbol{\beta}, \tau^2, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim \begin{cases} N\left(0, \tau^2\right) & \text{if } \eta_j \in \left(\max_{i\in S_k}\left\{\gamma_i^*\right\}, \min_{i\in S_k}\left\{\alpha_i^*\right\}\right) \\ 0 & \text{otherwise} \end{cases}$$

- The conditional posterior distribution of $\mathbf{U}$ and $\mathbf{V}$:

$$\begin{aligned} \pi_k(U_i|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{V}, A, \mathbf{y}) &\propto I\left[u_i < \exp\left\{y_i(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right\}\right] \\ \pi_k(V_i|\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta}, \mathbf{U}, A, \mathbf{y}) &\propto e^{-v_i} I\left[v_i > \exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right] \end{aligned}$$

A.3.2 Metropolis-Hastings

Let $Z_i \equiv \log(Y_i)$, then Z_i is a linear mixed Dirichlet model (LMDM). Details about the behavior and MCMC methods for LMDM are discussed in Kyung *et al.* (2008). For this model,

- the conditional posterior distribution of β:

$$\beta|\sigma^2, \tau^2, \eta, A, \mathbf{y} \sim N_p\left(\left(\frac{1}{d^*}I + X'X\right)^{-1} X'(\mathbf{Z} - A\eta), \sigma^2\left(\frac{1}{d^*}I + X'X\right)^{-1}\right). \tag{21}$$

- the conditional posterior distribution of η:

$$\eta|\beta, \sigma^2, \tau^2, \mathbf{U}, A, \mathbf{y} \sim N_k\left(\frac{1}{\sigma^2}\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1} A'(\mathbf{U} - X\beta), \left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1}\right). \tag{22}$$

Therefore, (21) is considered as a candidate density of $\boldsymbol{\beta}$ and (22) for $\boldsymbol{\eta}$.

The Metropolis-Hastings sampler of $(\boldsymbol{\beta}, \tau^2, \boldsymbol{\eta})$ follows.

- The conditional posterior distribution of $\boldsymbol{\beta}$:

$$\pi_k(\boldsymbol{\beta}|\tau^2, \boldsymbol{\eta}, A, \mathbf{y}) \propto e^{-\frac{1}{2d^*\sigma^2}|\boldsymbol{\beta}|^2} \prod_{i=1}^n e^{-\exp(\mathbf{X}_i\boldsymbol{\beta}+(\mathbf{A}\boldsymbol{\eta})_i)} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i}.$$

  Let

$$\pi_k^+(\boldsymbol{\beta}) \equiv e^{-\frac{1}{2d^*\sigma^2}|\boldsymbol{\beta}|^2} \prod_{i=1}^n e^{-\exp(\mathbf{X}_i\boldsymbol{\beta}+(\mathbf{A}\boldsymbol{\eta})_i)} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i}.$$

  For given $\boldsymbol{\beta}^{(t)}$,

  1. Generate $\boldsymbol{\beta}^* \sim N_p\left(\left(\frac{1}{d^*}I + X'X\right)^{-1} X'(\mathbf{Z} - A\eta), \sigma^2\left(\frac{1}{d^*}I + X'X\right)^{-1}\right)$.
  2. Take

$$\boldsymbol{\beta}^{(t+1)} = \begin{cases} \boldsymbol{\beta}^* & \text{with probability } \min\left\{\left(\frac{\pi^+(\boldsymbol{\beta}^*)}{\pi_k^+(\boldsymbol{\beta}^{(t)})}\frac{q(\boldsymbol{\beta}^{(t)})}{q(\boldsymbol{\beta}^*)}\right), 1\right\} \\ \boldsymbol{\beta}^{(t)} & \text{otherwise} \end{cases},$$

  where $q(\cdot)$ is a density of N_p distribution in (21), and recall that $\pi^+(\theta) = l(\theta)\pi(\theta)$.

- The conditional posterior distribution of τ^2:

$$\pi_k(\tau^2|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y}) \propto \left(\frac{1}{\tau^2}\right)^{k/2+a+1} e^{-\frac{1}{\tau^2}\left(\frac{1}{2}|\boldsymbol{\eta}|^2+b\right)}.$$

  Thus,

$$\tau^2|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}, A, \mathbf{y} \sim \text{Inverse Gamma}\left(\frac{k}{2} + a, \frac{1}{2}|\boldsymbol{\eta}|^2 + b\right).$$

- The conditional posterior distribution of $\boldsymbol{\eta}$:

$$\pi_k(\boldsymbol{\eta}|\boldsymbol{\beta}, \tau^2, A, \mathbf{y}) \propto \prod_{k=1}^{K} e^{-\frac{1}{2\tau^2}\eta_k^2} \prod_{i \in S_k} e^{-\exp(\mathbf{X}_i\boldsymbol{\beta}+(\mathbf{A}\boldsymbol{\eta})_i)} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i}.$$

For $j = 1, \ldots, k$, let

$$\begin{aligned} \pi_k^+(\eta_j) &\equiv e^{-\frac{1}{2\tau^2}\eta_j^2} \prod_{i \in S_j} e^{-\exp(\mathbf{X}_i\boldsymbol{\beta}+(\mathbf{A}\boldsymbol{\eta})_i)} \left[\exp(\mathbf{X}_i\boldsymbol{\beta} + (\mathbf{A}\boldsymbol{\eta})_i)\right]^{y_i} \\ &= e^{-\frac{1}{2\tau^2}\eta_j^2} \exp\left[\eta_j \sum_{i \in S_j} y_i - e^{\eta_j} \sum_{i \in S_j} e^{\mathbf{X}_i\boldsymbol{\beta}}\right]. \end{aligned}$$

For given $\boldsymbol{\eta}^{(t)}$,

1. Generate $\boldsymbol{\eta}^* \sim N_k\left(\frac{1}{\sigma^2}\left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1} A'(\mathbf{U} - X\beta), \left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}A'A\right)^{-1}\right)$.
2. Take

$$\boldsymbol{\eta}^{(t+1)} = \begin{cases} \boldsymbol{\eta}^* & \text{with probability } \min\left\{\left(\frac{\pi_k^+(\boldsymbol{\eta}^*)}{\pi_k^+(\boldsymbol{\eta}^{(t)})}\frac{q^*(\boldsymbol{\eta}^{(t)})}{q^*(\boldsymbol{\eta}^*)}\right), 1\right\} \\ \boldsymbol{\eta}^{(t)} & \text{otherwise} \end{cases},$$

where $q^*(\cdot)$ is a density of N_k distribution in (22).