Journal of the Korean Statistical Society (2002), **31**: 1, pp 93–107

Bayesian Parameter Estimation and Variable Selection in Random Effects Generalised Linear Models for Count Data†

Man-Suk Oh[1] and Tae Sung Park[2]

ABSTRACT

Random effects generalised linear models are useful for analysing clustered count data in which responses are usually correlated. We propose a Bayesian approach to parameter estimation and variable selection in random effects generalised linear models for count data. A simple Gibbs sampling algorithm for parameter estimation is presented and a simple and efficient variable selection is done by using the Gibbs outputs. An illustrative example is provided.

Keywords: Correlated data, Markov chain Monte Carlo, density estimation, Bayes factor, repeated measurements.

1. Introduction

Generalised linear models (GLM) have been widely used to analyse data with continuous and discrete response variables, as described in McCullagh and Nelder (1989). The underlying assumption for the generalised linear models is that the responses are independent with distributions from the exponential family. In many applications, however, responses may be obtained in clusters and may not be independent. In longitudinal studies, for example, responses repeatedly observed from the same subject are correlated. In sampling surveys, responses from the members of the same family are likely to be correlated.

One way to incorporate the correlation among responses is to introduce correlation parameters in addition to the mean parameters. Liang and Zeger (1986)

†This paper was accomplished with Research Fund provided by Korean Research Foundation, Support for Faculty Research Abroad.

[1]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea (e-mail : msoh@mm.ewha.ac.kr)

[2]Department of Statistics, Seoul National University, Seoul 151-747, Korea

proposed an approach based on the generalised estimating equations which are derived without full specification of the joint distribution of observations from a subject. Only the likelihood for the marginal distributions and a working correlation matrix for the vector of repeated measurements from each subject need to be specified. Their approach is closely related to quasi-likelihood methods.

An alternative way to incorporate the correlation is to add random effects to fixed covariate terms in the mean structure in GLM, resulting in the random effects GLM or generalised linear mixed models. Especially, a log-linear model with random effects have been proposed for correlated count data (Breslow, 1984; Tsutakawa, 1988).

In this paper, we consider a Bayesian analysis of the random effects GLM for count data. Bayesian approach is flexible in the sense that it can use important prior knowledges and the classical maximum likelihood estimates can be obtained from the posterior modes when constant priors are used. However, Bayesian analysis of random effects GLM for count data has not been widely used mainly because of computational difficulties.

Recently Markov chain Monte Carlo (MCMC) methods have drawn great attention as a tool for solving computational problems in various statistical applications. In particular, the Gibbs sampling algorithm described by Gelfand and Smith (1990) has been widely used since it is very easy to implement if the full conditional distributions of components of the parameter are given in convenient forms for random generation. Moreover, several authors recently developed methods which utilize outputs from the Gibbs sampler to estimate posterior density functions (Chib, 1997; Oh, 1999) which can be used for simple and efficient variable selection, as will be shown in Section 5.

In the random effects GLM for count data, however, implementation of the Gibbs sampling algorithm is not straightforward since the posterior distributions of parameters are not given in convenient forms by lack of conjugacy among the distributions of count data, the random effects, and the prior (see Section 2). To overcome the difficulty and hence utilize the advantages of the Gibbs sampler for posterior inference and variable selection, we propose to introduce some auxiliary variables into the model so that they can lead to convenient full conditional density function of each component of parameters, which is essential for easy implementation of the Gibbs sampler. Previous works on using auxiliary variables to relieve the complexity in full conditional distributions in the Gibbs sampler are Albert and Chib (1993) and Oh (1998), among others.

Relavant MCMC methods for random effects GLM for count data are Zeger and Karim (1991) and Chib *et al.* (1998). Zeger and Karim (1991) suggested a Gibbs sampling algorithm in which a rejection sampling is used for generating components of parameters. Their method is very general in that it can be applied to any type of data. However, the rejection sampling with a normal sample generating density may be very inefficient when the normal approximation is not good. Recently Chib *et al.* (1998) suggested a Metropolis-Hastings algorithm for random effects generalised linear models for count data. One may estimate post density functions from the Metropolis-Hastings outputs (Chen, 1995). However, it requires some correction factors (importance weights) which may have a large variation and cause inefficiency in estimating the posterior densities.

The paper is organized as follows. Section 2 describes the random effects generalised linear models. Section 3 introduces the auxiliary variables. The full conditional posterior distributions of components of parameters are derived in Section 4. A simple and efficient variable selection scheme utilizing the Gibbs outputs is introduced in Section 5. An illustrative example is given in Section 6 and some concluding remarks are given in Section 7.

2. Random Effects Generalized Linear Models

Let y_{ij} be a count response data and x_{ij} be the vector of p predictors, for the jth subject in cluster i, $j = 1, ..., n_i$, $i = 1, .., n$. Here, the observations $y_{i1}, ..., y_{i,n_i}$ in cluster i are likely to be correlated.

A random effects GLM is defined as follows. Given the p-dimensional fixed effect β and the q-dimensional random effect b_i, the response y_{ij} follows an exponential distribution

$$f(y_{ij}|\theta_{ij}) = \exp[(y_{ij}\theta_{ij} - a(\theta_{ij}) + Q(y_{ij}))/\phi].$$

We consider the case where the scale parameter ϕ is 1 and the link function is the canonical link function. Then a random effects GLM is given by

$$\theta_{ij} = \eta_{ij} = g(\beta, b_i) = x'_{ij}\beta + z'_{ij}b_i,$$

where z_{ij} is a q-dimensional covariate vector. In addition, the density function of y_{ij} can be written as

$$f(y_{ij}|\beta, b_i) = \exp[y_{ij} \cdot g(\beta, b_i) + S(\beta, b_i) + Q(y_{ij})].$$

Since we are interested in count data, we mainly focus on the Poisson response or binary response in which it is reasonable to assume that the scale parameter ϕ

is 1. In addition, the canonical link functions are most commonly used for these responses.

The random effects b_i's are assumed to be independent and identically distributed a certain parametric distribution. The most common choice for the distribution of b_i is a multivariate normal distribution with mean η and covariance matrix D, i.e., $N(\eta, D)$. We used a nonzero mean for b_i to avoid nonidentifiability in MCMC, as suggested in Chib *et al.* (1998).

In the above random effects GLM, the parameters of interest are (β, η, D) and the likelihood function of (β, η, D) given data $\{x_{ij}, y_{ij}\}$ is given by

$$l(\beta, \eta, D|data) \propto \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij}|\beta, b_i) \cdot |D|^{-1/2} \exp[-\frac{1}{2}(b_i - \eta)' D^{-1}(b_i - \eta)] \; db_i. \tag{2.1}$$

Combining the likelihood and a prior $\pi(\beta, \eta, D)$, one obtains a posterior density function

$$\pi(\beta, \eta, D|data) = \frac{l(\beta, \eta, D|data)\pi(\beta, \eta, D)}{\int l(\beta, \eta, D|data)\pi(\beta, \eta, D) d\beta d\eta dD}, \tag{2.2}$$

and any posterior inference is based on the posterior density.

For count data, obviously the posterior density is not given in a closed form and a numerical approximation is required. We propose to use the Gibbs sampling algorithm for posterior inference due to its advantages mentioned in Section 1.

3. Auxiliary Variables

The Gibbs sampler iteratively generates samples from full conditional distributions of components of parameters. Thus, easy random generation from the full conditional posterior distributions is essential for the Gibbs sampler. For count data, however, due to the non-conjugacy among the conditional density $f(y_{ij}|\beta, b_i)$, the mixing density $f(b_i|\eta, D)$ of b_i, and the prior $\pi(\beta, \eta, D)$, the full conditional posterior distributions of components of the parameters are very complicated and direct implementation of the Gibbs sampler is prohibited. Especially, the nonconjugacy between $f(y_{ij}|\beta, b_i)$ and $f(b_i|\eta, D)$ makes even the likelihood $l(\beta, \eta, D|data)$ analytically intractable. To avoid this nonconjugacy problem, we introduce normal auxiliary variables into the model, extending the ideas of Albert and Chib (1993) and Oh (1997), for random effects GLM. With these extra auxiliary variables, the full conditional distributions of components of the parameters are given in standard forms, so that the Gibbs algorithm can be easily applied.

Given $(\beta,\ b_i)$, let W_{ij} be a random variable following $N(x'_{ij}\beta + z'_{ij}b_i, 1)$ distribution. Define

$$Y_{ij} = y_{ij} \text{ if } \Phi^{-1}[F(y_{ij}-1|\beta,b_i)] < W_{ij} - x'_{ij}\beta - z'_{ij}b_i \leq \Phi^{-1}[F(y_{ij}|\beta,b_i)],$$

where Φ is the standard normal cdf and F is the cdf of Y_{ij} which depends on (β, b_i). Then the joint density function of (Y_{ij}, W_{ij}) given $(\beta,\ b_i)$ is

$$\begin{aligned} & f(y_{ij}, w_{ij}|\beta, b_i) \\ & \quad = \phi(w_{ij}|x'_{ij}\beta + z'_{ij}b_i, 1) \\ & \qquad \times I(\Phi^{-1}[F(y_{ij}-1|\beta,b_i)] < W_{ij} - x'_{ij}\beta - z'_{ij}b_i \leq \Phi^{-1}[F(y_{ij}|\beta,b_i)]), \end{aligned}$$

where $\phi(\cdot|\mu,\sigma^2)$ is the density function of $N(\mu,\sigma^2)$ and I is the indicator function. It can be easily shown that the marginal density function of Y_{ij} derived from the joint density $f(y_{ij}, w_{ij}|\beta, b_i)$ correctly gives the density $f(y_{ij}|\beta, b_i)$ corresponding to the cdf $F(y_{ij}|\beta, b_i)$.

With the extra auxiliary variables W_{ij} in the model, we may consider $\theta = (\beta, \eta, D, \{b_i\},$
$\{W_{ij}\})$ as new unknown parameters in the model. Now, the posterior density function of θ is given as

$$\begin{aligned} \pi(\theta|data) & \propto \prod_{i=1}^{n}\{\prod_{j=1}^{n_i} f(y_{ij}, w_{ij}|\beta, b_i) \cdot |D|^{-1/2}\exp[-\frac{1}{2}(b_i-\eta)'D^{-1}(b_i-\eta)]\} \\ & \quad \times\pi(\beta,\eta,D) \\ & \propto \exp[-\frac{1}{2}\{\sum_i\sum_j(w_{ij} - x'_{ij}\beta - z'_{ij}b_i)^2 + \sum_i(b_i-\eta)'D^{-1}(b_i-\eta)\}] \\ & \quad \times|D|^{-n/2}\ \pi(\beta,\eta,D)\ \prod_i\prod_j I(A_{ij}(\beta, b_i, w_{ij})), \end{aligned}$$

where

$$\begin{aligned} & A_{ij}(\beta, b_i, w_{ij}) \\ & \quad = \{(\beta, b_i, w_{ij}); \Phi^{-1}F(y_{ij}-1|\beta,b_i) < w_{ij} - x'_{ij}\beta - z'_{ij}b_i \leq \Phi^{-1}F(y_{ij}|\beta,b_i)\}. \end{aligned}$$

In the next section, we derive full conditional distributions of W_{ij}, b_i, β, η and D.

4. Full Conditional Posterior Distributions

Suppose that $\pi(\beta, \eta, D) = \pi(\beta)\pi(\eta)\pi(D)$, where $\pi(\beta)$ and $\pi(\eta)$ are densities of $N(\beta_0, \Sigma_0)$ and $N(\eta_0, \Psi_0)$, respectively, and $\pi(D)$ is the density of $IW(m, E)$, the inverted Wishart distribution with parameters m and E (Press, 1988). Then the posterior density becomes

$$\begin{aligned}\pi(\theta|data) \propto\ & \exp[-\frac{1}{2}\{\sum_i \sum_j (w_{ij} - x'_{ij}\beta - z'_{ij}b_i)^2 + \sum_i (b_i - \eta)' D^{-1}(b_i - \eta) \\ & +(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0) + (\eta - \eta_0)'\Psi_0^{-1}(\eta - \eta_0) + tr(D^{-1}E)\}] \\ & \times |D|^{-(n+m+q+1)/2} \prod_i \prod_j I(A_{ij}(\beta, b_i, w_{ij})), \qquad (4.1)\end{aligned}$$

where tr denotes the trace of a matrix. From this posterior kernel, it can be easily seen that the full conditional density of η and D are given as a normal and an inverted Wishart density, respectively. In addition, if we ignore the restriction in the indicator function, then the logarithm of the posterior is a quadratic function of w_{ij}, b_i, β. Hence, the full conditional posterior distributions of w_{ij}, b_i, β are all restricted normal distributions. Now we describe each full conditional posterior distribution in detail.

Let $X_i = (x_{i1}, ..., x_{i,n_i})'$, $Z_i = (z_{i1}, ..., z_{i,n_i})'$ and $W_i = (w_{i1}, ..., w_{i,n_i})'$. We also let $A_{ij}(\beta|b_i, w_{ij})$ denote the subregion of β such that, for given (b_i, w_{ij}), (β, b_i, w_{ij}) is in $A_{ij}(\beta, b_i, w_{ij})$. In the same manner, we denote the subregion of b_i and W_{ij} in $A_{ij}(\beta, b_i, w_{ij})$, given others, by $A_{ij}(b_i|\beta, w_{ij})$ and $A_{ij}(W_{ij}|\beta, b_i)$, respectively. In what follows, we will suppress the data in the posterior distribution and let $[\beta|others]$, $[b_i|others]$, $[W_{ij}|others]$, $[\eta|others]$, $[D|others]$, denote respectively the full conditional posterior distributions of β, b_i, W_{ij}, η and D, given all the other unknown variables and data.

From the posterior kernel (4.1), the following full conditional posterior distributions are obtained:

$$\begin{aligned}& [\beta|others] \sim N(\mu_\beta, \Sigma_\beta)\ I(\cap_{i=1}^n \cap_{j=1}^{n_i} A_{ij}(\beta|b_i, w_{ij})), \\ & [b_i|others] \sim N(\mu_{b_i}, \Sigma_{b_i})\ I(\cap_{j=1}^{n_i} A_{ij}(b_i|\beta, w_{ij})), \\ & [W_{ij}|others] \sim N(x'_{ij}\beta + z'_{ij}b_i, 1)\ I(A_{ij}(W_{ij}|\beta, b_i)), \\ & [\eta|others] \sim N(\mu_\eta, \Sigma_\eta), \\ & [D|others] \sim IW(n + m, E + \sum_{i=1}^n (b_i - \eta)(b_i - \eta)'), \qquad (4.2)\end{aligned}$$

where

$$\begin{aligned}
\mu_\beta &= \Sigma_\beta\left(\sum_i X_i'(W_i - Z_i b_i) + \Sigma_0^{-1}\beta_0\right), & \Sigma_\beta &= \left(\sum_i X_i' X_i + \Sigma_0^{-1}\right)^{-1}, \\
\mu_{b_i} &= \Sigma_{b_i}(Z_i'(W_i - X_i\beta) + D^{-1}\eta), & \Sigma_{b_i} &= (Z_i' Z_i + D^{-1})^{-1}, \\
\mu_\eta &= \Sigma_\eta(D^{-1}\sum_i b_i + \Psi_0^{-1}\eta_0), & \Sigma_\eta &= (nD^{-1} + \Psi_0^{-1})^{-1}
\end{aligned} \tag{4.3}$$

and $\cap$ denotes the intersection of intervals.

There is no restriction in the full conditional posterior distribution of η and D, hence generation of samples of η and D is straightforward. One may use the algorithm of Odell and Feiveson (1966) for sample generation from a Wishart distribution. In addition, since $A_{ij}(W_{ij}|\beta, b_i)$ can be expressed as

$$\begin{aligned}
A_{ij}(W_{ij}|\beta, b_i) &= \{W_{ij}\ ; \Phi^{-1}F(y_{ij} - 1|\beta, b_i) + x_{ij}'\beta + z_{ij}'b_i \\
&\qquad < W_{ij} \le \Phi^{-1}F(y_{ij}|\beta, b_i) + x_{ij}'\beta + z_{ij}'b_i\},
\end{aligned}$$

the full conditional posterior distribution of W_{ij} is a univariate normal distribution restricted to a fixed interval, from which sample generation can be done very easily by using the cdf inverse method (Devroye, 1986), the mixed rejection algorithm (Geweke, 1991), or a simple accept-rejection method (Rubinstein, 1988).

On the other hand, the full conditional posterior distributions of β and b_i are restricted multivariate normals with highly complicated forms of restrictions and hence generation of samples of β and b_i is not easy. The Hit-Miss algorithm described in Rubinstein (1988) may be used, which keeps generating candidate samples from the unrestricted normal distribution until one obtains a sample belonging to the region. However, it is well known that the Hit-Miss algorithm can be very inefficient when the restricted region has a very small probability, which is often the case in a high dimensional space. Thus, the Hit-Miss algorithm would be inappropriate for the sample generation of β and b_i.

Because the Gibbs sampler iteratively generates samples from the full conditional distributions, a simple and efficient method of generating samples from the full conditional distributions of β and b_i is desirable. For this, we consider a simple and accurate approximation to the restriction in the full conditional distributions of β and b_i.

The complexity of the restricted regions of β and b_i arises mainly from $\Phi^{-1}F(y_{ij}|\beta, b_i)$, for which a simple closed expression is not known for count data

y_{ij}. However, for some distributions for count data such as Poisson and Binomial, the cdf F can be very accurately approximated by the standard normal cdf, with appropriate transformations. With this approximation, the complicated $\Phi^{-1}F$ can be greatly simplified, resulting in a simple approximation to the restriction. We illustrate this for the Poisson data case in the rest of this section. Other distributions can be handled in a similar way.

When y follows a Poisson distribution with mean λ, as shown in Johnson and Kotz (1972, Ch.4) and Oh and Lim (2001), the cdf $F(y|\lambda)$ of y is almost identical to

$$F^*(y|\lambda) = \Phi[-3\sqrt{y+1}\left[(\frac{\lambda}{y+1})^{1/3} - 1 + \frac{1}{9(y+1)}\right]].$$

Thus, we may replace F by F^* and hence $\Phi^{-1}F(y|\lambda)$ by

$$\Phi^{-1}F^*(y|\lambda) \quad = \quad -3\sqrt{y+1}\left[\frac{\lambda^{1/3}}{(y+1)^{1/3}} - 1 + \frac{1}{9(y+1)}\right].$$

Therefore, the restriction $A_{ij}(\beta, b_i, w_{ij})$ can be replaced by

$$\begin{aligned} & A^*_{ij}(\beta, b_i, w_{ij}) \\ & = \{(\beta, b_i, w_{ij}); a(y_{ij}-1, x'_{ij}\beta + z'_{ij}b_i, w_{ij}) < 0 \le a(y_{ij}, x'_{ij}\beta + z'_{ij}b_i, w_{ij})\}, \end{aligned} \tag{4.4}$$

where

$$a(y_{ij}, x'_{ij}\beta + z'_{ij}b_i, w_{ij}) = \Phi^{-1}F^*(y_{ij}|e^{x'_{ij}\beta + z'_{ij}b_i}) - w_{ij} + x'_{ij}\beta + z'_{ij}b_i. \tag{4.5}$$

Now it remains to solve the inequalities

$$a(y_{ij}-1, \delta_{ij}, w_{ij}) < 0 \le a(y_{ij}, \delta_{ij}, w_{ij}), \tag{4.6}$$

for δ_{ij} given y_{ij} and w_{ij}, where $\delta_{ij} = x'_{ij}\beta + z'_{ij}b_i$. Using the fact that a is increasing in y_{ij} and concave in δ_{ij} and that there exist two distinct solutions of δ_{ij} for $a(y_{ij}, \delta_{ij}, w_{ij}) = 0$, it can be shown that (4.6) is equivalent to $\{c_1(w_{ij}) < \delta_{ij} < c_2(w_{ij})\}$ or

$$\{c_1(w_{ij}) < \delta_{ij} < d_1(w_{ij})\} \cup \{d_2(w_{ij}) < \delta_{ij} < c_2(w_{ij})\},$$

where $c_1(w_{ij})$ and $c_2(w_{ij})$ are two distinct solutions of $a(y_{ij}, \delta, w_{ij}) = 0$, $d_1(w_{ij})$ and $d_2(w_{ij})$ are two distinct solutions of $a(y_{ij}-1, \delta, w_{ij}) = 0$, and $\cup$ denotes the union of intervals (Oh and Lim, 2001). Thus, given w_{ij}, the region of $\delta_{ij} = x'_{ij}\beta + z'_{ij}b_i$ satisfying the restriction A^*_{ij} is given as a fixed interval, hence, given w_{ij} and all the other parameters, the region of each element of β and b_i is given as a fixed interval. From (4.2) with A_{ij} replaced by A^*_{ij}, the full conditional distributions of elements of β and b_i are given as restricted normal distributions.

5. Variable Selection

Let $M_0 : \beta \neq 0$ and $M_i : \beta_{(i)} = 0$ where $\beta_{(i)}$ is the i-th subset of β among all possible $2^p - 1$ subsets.

Bayesian variable selection is based on the posterior probability

$$\pi(M_i|data) = \frac{\int p(data|\beta_{-(i)}, M_i)\pi(\beta_{-(i)}|M_i)d\beta_{-(i)}p(M_i)}{\sum_{i=0}^{2^p-1} \int p(data|\beta_{-(i)}, M_i)\pi(\beta_{-(i)}|M_i)d\beta_{-(i)}p(M_i)},$$

where $\beta_{-(i)}$ is the set of elements of β other then $\beta_{(i)}$, $p(M_i)$ is the prior probability of model M_i, and $\pi(\beta_{-(i)}|M_i)$ is the prior density of $\beta_{-(i)}$ under M_i. Suppose that $\pi(\beta_{-(i)}|M_i) = \pi(\beta_{-(i)}|\beta_{(i)} = 0, M_0)$ and $p(M_i)$ are all equal,

$$\begin{aligned}
\frac{\pi(M_i|data)}{\pi(M_0|data)} &\propto \int p(data|\beta_{-(i)}, M_i)\pi(\beta_{-(i)}|M_i)d\beta_{-(i)} \\
&= \int p(data|\beta_{-(i)}, \beta_{(i)} = 0, M_0)\pi(\beta_{-(i)}|\beta_{(i)} = 0, M_0)d\beta_{-(i)} \\
&= \int \frac{p(data|\beta_{-(i)}, \beta_{(i)} = 0, M_0)\pi(\beta_{-(i)}, \beta_{(i)} = 0, M_0)}{\pi(\beta_{(i)} = 0|M_0)}d\beta_{-(i)} \\
&= \frac{\pi(\beta_{(i)} = 0|data, M_0)}{\pi(\beta_{(i)} = 0|M_0)} \\
&\equiv B_i,
\end{aligned}$$

where $\pi(\beta_{(i)} = 0|data, M_0)$ is the posterior density of $\beta_{(i)}$ at 0 under M_0.

Thus, the posterior model probabilities are given as

$$\begin{aligned}
P(M_0|data) &= (1 + \sum_{i=1}^{2^p-1} B_i)^{-1}, \\
P(M_i|data) &= B_i \cdot P(M_0|data), \ i = 1, .., 2^p - 1.
\end{aligned}$$

Once the posterior model probabilities are obtained, one may choose the model with the highest posterior probability or perform model averaging, i.e., combine information on β from models with significant posterior probabilities. (Kass and Raftery, 1995; Hoeting *et al.*, 1999)

Now, the variable selection problem becomes the problem of estimating marginal posterior density functions of all possible subsets of β. Estimating each marginal density function from a different MCMC algorithm, as suggested in Chib (1995) and Chen (1994), is not practical here since 2^p increases very rapidly as p.

Oh (1999) suggested a method which can simultaneously estimate all marginal posterior density functions by using only one set of posterior samples. Especially when all the full conditional posterior density functions of components of parameters are given, the method can be applied without any tailoring, and all marginal density functions can be estimated *simutaneously during the simulation.* For instance, from Oh (1999),

$$\begin{aligned}
&\pi(\beta_1 = \beta_2 = 0|data, M_0)\\
&= E[\pi(\beta_2 = 0|\beta_1 = 0, \beta_3, .., \beta_p, \{w_{ij}\}, \{b_i\}, data, M_0)\\
&\qquad \times \pi(\beta_1 = 0|\beta_2, .., \beta_p, \{w_{ij}\}, \{b_i\}, data, M_0)],\\
&\pi(\beta_1 = \beta_2 = \beta_3 = 0|data, M_0)\\
&= E[\pi(\beta_3 = 0|\beta_1 = \beta_2 = 0, \beta_4, .., \beta_p, \{w_{ij}\}, \{b_i\}, data, M_0)\\
&\qquad \times \pi(\beta_2 = 0|\beta_1 = 0, \beta_3, .., \beta_p, \{w_{ij}\}, \{b_i\}, data, M_0)\\
&\qquad \times \pi(\beta_1 = 0|\beta_2, .., \beta_p, \{w_{ij}\}, \{b_i\}, data, M_0)],
\end{aligned}$$

where the expectations are taken with respect to the joint posterior distribution of β, $\{w_{ij}\}$ *and* $\{b_i\}$.

As shown in Section 4, the full conditional density function of each element of β and b_i is given as a univariate restricted normal density, hence estimation of the posterior density functions is strightforward. Note also that it does not require any additional sampling since all the marginal densities are expressed as expectations with respect to the same joint posterior distribution, which is a great advantage of the method for the purpose of variable selection.

6. An Illustrative Example

We applied the proposed algorithm to the hospital visit data given by Karim and Zeger (1989). The numbers of hospital visits made by a group of 73 children were observed during four different time periods. The sex of each child, smoking status of the child's mother during pregnancy, and the child's age in months were recorded. We assumed a Poisson distribution with mean $\lambda_{ij} = \exp[x_{ij}'\beta + z_{ij}b_i]$ for the response, the number of hospital visits, and assumed $b_i \sim N(\eta, D)$. For the covariates x_{ij}, we included the indicators of second, third, fourth visit, sex and smoke, and logarithm of the age divided by 12. We let $z_{ij} = 1$ for all i, j and so b_i is univariate. We did not include the intercept term in x_{ij} but allow nonzero mean for b_i instead since Chib *et al.* (1998) pointed out that common terms in x_{ij} and z_{ij} can cause a nonidentifiability problem in Markov chain Monte Carlo methods.

For prior distributions for the parameters β, η and D, we assumed $\pi(\beta, \eta, D) =$

TABLE 6.1 *Posterior means and standard deviations of parameters*

	Estimate	*SD*		*Estimate*	*SD*
β_1	−0.43558	0.15726	β_5	0.32496	0.25436
β_2	−0.30621	0.15096	β_6	0.12603	0.18293
β_3	−1.01192	0.18768	η	−0.27552	0.31730
β_4	−0.10977	0.24338	D	0.91887	0.47456

$\pi(\beta)\pi(\eta)\pi(D)$ and chose the following vague priors for β, η, D :

$$\beta \sim N(\hat{\beta}, 30I), \quad \eta \sim N(0,1), \quad D \sim IG(5,5),$$

where $\hat{\beta}$ is the MLE of β obtained from the generalised linear model without random effect, I is the identity matrix, and $IG(a,b)$ represents the inverse Gamma distribution which corresponds to one dimensional inverted Wishart distribution with parameters $2a$ and $4b$.

As starting values in the Gibbs sampling algorithm, we used $\beta = \hat{\beta}$, $b_i = 0$, $\eta = 0$, $D = 1$ and $w_{ij} = 1$. We applied the proposed Gibbs sampling algorithm by using the full conditional posteriors given in Section 4 and ran 300,000 iterations after 50,000 iterations for burn-in. Time sequence plots of the unknown parameters, given in Figure 6.1, suggest that the convergence has been achieved.

Table 6.1 presents posterior estimates of β, η and D from the Gibbs sampler. It can be observed that the last three elements of β are not significant in the sense that their 95% credible intervals contain 0.

We estimated the marginal posterior density functions at zeros for all possible subsets of β and then computed the posterior probabilities of the models. Models with posterior probability greater than 0.05 are presented in Table 6.2 in descending order of their posterior probabilities. It can be observed that the model with only β_1 and β_3, i.e., with $\beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$ has the highest posterior probability and hence this model can be selected as the best model. However, the highest probability is only 0.3113 and the other five models listed in Table 6.2 have probability of roughly 0.1. Thus, it might be reasonable to consider model averaging for inferential purposes.

7. Concluding Remarks

In this paper, we have developed a Gibbs sampling algorithm for parameter estimation and variable selection in random effects generalised linear models

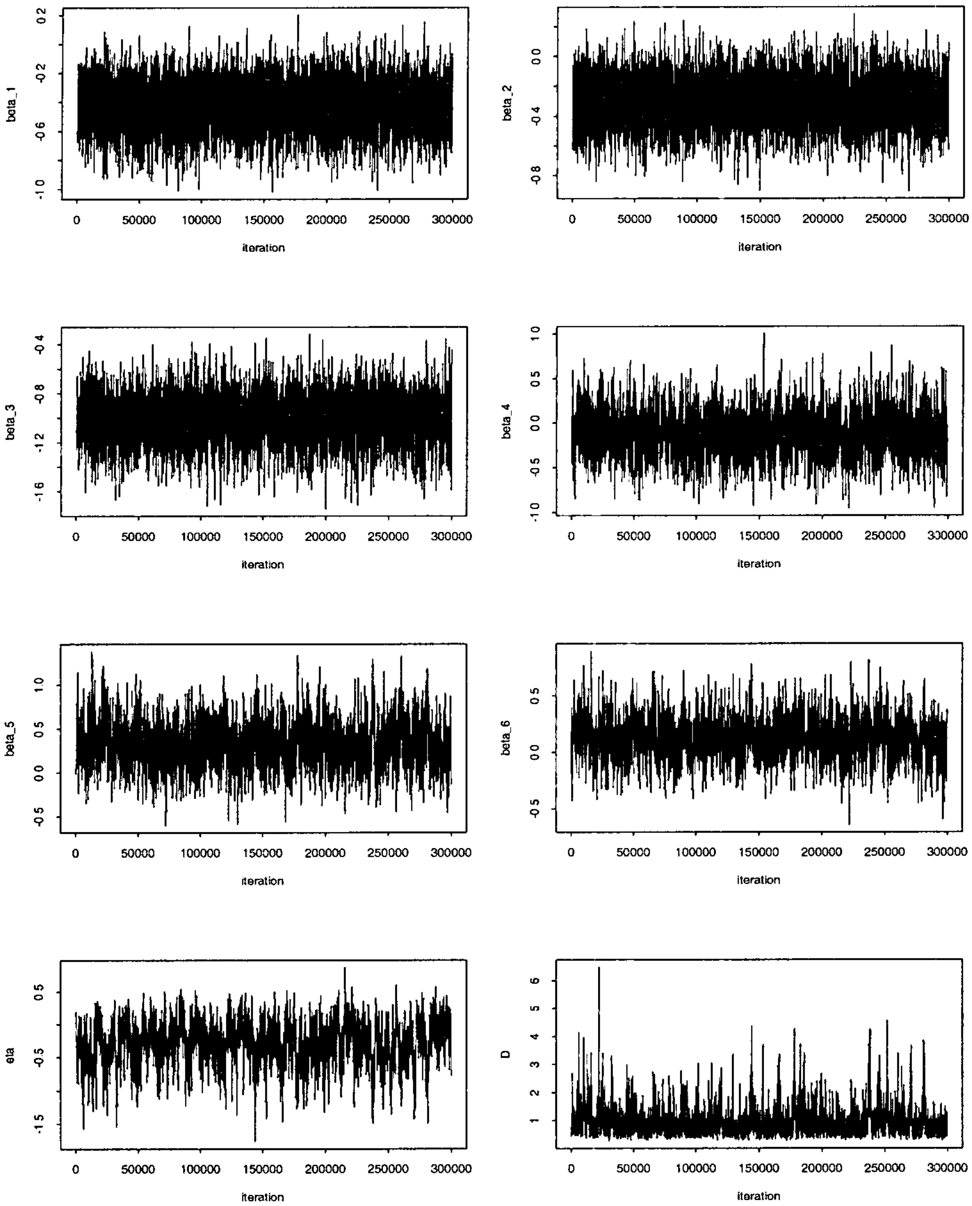

FIGURE 6.1 *Time sequence plots of parameters*

TABLE 6.2 *Models with significant posterior probabilities*

Model	Posterior Probability	Model	Posterior Probability
(β_1, β_3)	0.3113	$(\beta_1, \beta_2, \beta_3)$	0.1042
(β_3, β_5)	0.1203	$(\beta_1, \beta_3, \beta_4)$	0.0830
(β_3, β_6)	0.1196	$(\beta_1, \beta_3, \beta_5)$	0.0759

for count data. The key feature of the method is that it introduces some normal auxiliary variables into the model, so that all the full conditional posterior distributions of the fixed and the random coefficients in the model are given as restricted normal distributions. The complicated restrictions in the normal distributions due to the non-conjugacy among the distributions of the count data, the random effect, and the prior are simplified by using simple and accurate approximations to the restrictions.

Obviously, the algorithm can also be used for the generalised linear model without random effects. When there is no random effects in the model, one only need to skip generating b_i, η and D in the algorithm.

When count data follow Poisson or Binomial distribution, the cdf can be very accurately approximated by the standard normal cdf with appropriate transformations. In other cases, one may use a piecewise linear combinations of the standard normal cdf as in Oh (1997).

REFERENCES

Albert, J. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data: A Gibbs sampling approach", *Journal of the American Statistical Association*, **88**, 669-679.

Breslow, N. E. (1984). "Extra-Poisson variation in log-linear models", *Applied Statistics*, **33**, 38-44.

Chen, M-H. (1994). "Importance-weighted marginal Bayesian posterior density estimation", *Journal of the American Statistical Association*, **89**, 818-824.

Chib, S. (1995). "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, **90**, 1313-1321.

Chib, S. , Greenberg, E. and Winkelmann, R. (1998). "Posterior simulation and

Bayes factors in panel count data models", *Journal of Econometrics*, **86**, 33-54.

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

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, **85**, 398-409.

Geweke, M. (1991). "Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints", *Computing Science and Statistics (Proceedings of the 23rd Symposium on the Interface)*.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). "Bayesian model averaging: A tutorial", *Statistical Science*, **14**, 382-412.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics*, John & Wiley, New York.

Karim, M. R. and Zeger, S. (1989), GEE1 PC SAS, *Technical Report #674*, Dept. of Biostatistics, The Johns Hopkins University, Baltimore, MD.

Kass, R. E. and Raftery, A. E. (1990). "Bayes factors", *Journal of the American Statistical Association*, **90**, 773-795.

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

McCullagh, P. and Nelder, J. A. (1989). *Generalised Linear Models* (2nd ed.), Chapman & Hall, London.

Odell, P. L. and Feiveson, A. H. (1966). "A numerical procedure to generate a sample covariance matrix", *Journal of the American Statistical Association*, **61**, 199-203.

Oh, M-S. (1998). "Gibbs sampling approach to Bayesian analysis of generalised linear models for binary data", *Computational Statistics*, **12**, 431-446.

Oh, M-S. (1999). "Estimation of posterior density functions from a posterior sample", *Computational Statistics and Data Analysis*, **29**, 411-427.

Oh, M-S. and Lim, Y. B. (1999). "Bayesian analysis of time series Poisson count data", *Journal of Applied Statistics*, **28**, 259-271.

Press, J. S. (1988). *Bayesian Statistics*, Wiley, New York.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley & Sons, New York.

Tsutakawa, R. K. (1988). "Mixed models for analyzing geographic variability in mortality rates", *Journal of the American Statistical Association*, **83**, 37-42.

Zeger, S. L. and Karim, M. R. (1991). "Generalised linear models with random effects; a Gibbs sampling approach", *Journal of the American Statistical Association*, **86**, 79-86.