# Bayesian Logistic Regression Model Choice via Laplace-Metropolis Algorithm

**Farzad Eskandari**[1], **M. Reza Meshkani**[2]

[1]Department of Statistics, Allameh Tabatabai University, Iran.
(f-eskandari@cc.sbu.ac.ir)
[2]Department of Statistics, Shahid Beheshti University, Iran.
(m-meshkani@sbu.ac.ir)

**Abstract.** Following a Bayesian statistical inference paradigm, we provide an alternative methodology for analyzing a multivariate logistic regression. We use a multivariate normal prior in the Bayesian analysis. We present a unique Bayes estimator associated with a prior which is admissible. The Bayes estimators of the coefficients of the model are obtained via MCMC methods. The proposed procedure is illustrated by analyzing a data set which has previously been analyzed by various authors. It is shown that our model is more precise and computationally less taxing.

## 1   Introduction

In most areas of scientific research where statistical models are used, non-linear models are essential tools for analysing the results. They are used when one has a discrete and/or non-linear response. One

---

such important model is the logistic model which is used to explore the effects of some covariates, discrete and/or continuous independent variables, on a discrete response. Whenever both the response and covariates are discrete, one has a contingency table. Two important areas of current interest in which such models appear are logistic regression with covariate measurement error and random effects models for binary outcomes. These topics can be dealt with either by the frequentist or the Bayesian methods. From a Bayesian point of view, for example Dellaportas and smith(1993) demonstrated an adaptive-rejection sampling(ARS) from the univariate conditional prior densities of the coefficients in the logistic regression model. Gamerman(1997) uses a normal approximation to the posterior density of the regression coefficients using iterative-reweighted-least-squares(IRLS). The Gamerman's approach requires Metropolis-Hastings updates and hence data dependent accept-reject steps. Chen and Dey(1998) described a Bayesian logistic regression model based on the scale mixture method. They use a Bayesian model selection for multinomial logistic regression by using a scale mixture of normals representation for the noise process. Madigan et al. (2005) consider a problem similar to ours in the context of author identification, using both normal and Laplace priors for regression coefficients. Wang and Kapat (2006) present four types of priors, but they also recommend the use of normal and Laplace priors.

This article presents a Bayesian analysis for contingency tables where a logistic model is used. We focus on Laplace-Metropolis algorithm to approximate the marginal distribution. It also provides an algorithm for implementation of this method. We consider both normal and non-informative priors. When the researcher has enough prior information, he can use the normal specification. However, if he does not have or is not willing to specify such a prior, he may use the non-informative prior. Both cases have been explored, here. The proposed method is illustrated by a real example, which has previously been used by Schull(1958) and later by Ntzoufras et al.(1999) to illustrate their proposed methods. Originally, Schull(1958) studied the pregnancy outcome in three districts of Shizuoka city, Japan, according to the degree of consanguinity between the parents. In his study, Death ( categorized as abortion, stillbirth, death in less than 12 months, death in 13-60 months, survived) is considered as a multinomial response variable, and Residence( rural district, intermediate district and urban district) and Consanguinity ( no relation,

2nd cousins, $1\frac{1}{2}$ cousins and 1st cousins) are covariates. Ntzoufras et al. (1999) developed Metropolis-Hastings algorithm for exact inference on binomial and multinomial logistic regression models based on repeated categorical response.

In this paper, we will study this problem by the subjective Bayesian theory via MCMC, considering the logistic regression analysis for a multinomial response variable. In Section 2, the model is presented and for the logistic regression model the Bayes estimates of parameters are obtained.In Section 3, a real data set is analyzed. A small scale simulation in section 4 is reported.

## 2   The Model

Let **Y** be a multinomial response variable with K+1 categories which are denoted by $0, 1, \cdots, K$. Supose for $i = 1, 2, \cdots, I$, the independent vectors $\mathbf{Y}_i$ are distributed multinomially with probability vectors $\mathbf{p}_i = (p_{i0}, \cdots, p_{iK})$, that is, $Y_i' \sim Mult(N_i, \mathbf{p}_i)$, with fixed $N_i$. The counts matrix $Y = (y_{ij})$, for $i = 1, 2, \cdots, I$ and $j = 0, 1, \cdots, K$ contains the data. Columns of **Y** are denoted by $(\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_K)$, where $\mathbf{y}_0 = \mathbf{N} - \sum_{j=1}^{K} \mathbf{y}_j$, with $\mathbf{N} = (N_1, \cdots, N_I)'$. The likelihood function is

$$L(\mathbf{P}|\mathbf{Y} = \mathbf{y}) = \Big(\prod_{i=1}^{I} \frac{N_i!}{\prod_{j=0}^{K} y_{ij}!}\Big) exp\Big(\sum_{i=1}^{I} \sum_{j=0}^{K} y_{ij} log(p_{ij})\Big), \qquad (2.1)$$

where $\mathbf{P} = (p_{ij})$, $0 < p_{ij} < 1$ and $\sum_{j=0}^{K} p_{ij} = 1$, for each $i$. Now suppose for $j = 1, 2, \cdots, K$, $\boldsymbol{\beta}_j' = (\beta_{j0}, \beta_{j1}, \cdots, \beta_{jq})$ is the vector of regression parameters corresponding to the vector of $q$ covariates $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{iq})$. Using log odds $log(\frac{p_{ij}}{p_{i0}}) = \eta_{ij}$, the assumed structure for the logistic regression is

$$\eta_{ij} = \mathbf{x}_i' \beta_j, \qquad\qquad i = 1, 2, \cdots, I. \qquad\qquad (2.2)$$

or for $\boldsymbol{\eta}_j = (\eta_{1j}, \cdots, \eta_{Ij})$, we can write $\boldsymbol{\eta}_j = X\boldsymbol{\beta}_j$, with

$$X = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_I' \end{pmatrix}.$$

From (2.1) and (2.2), the canonical likelihood function for $\beta = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_K)$ is equal to:

$$L(\beta|\mathbf{Y} = \mathbf{y}) =$$
$$\left(\prod_i \frac{N_i!}{\prod_j y_{ij}!}\right) exp\left(\sum_i \sum_j y_{ij}\mathbf{x}_i'\beta_j - \sum_i N_i log(1 + \sum_j exp(\mathbf{x}_i'\beta_j))\right). \quad (2.3)$$

As with any statistical model, we must avoid overfitting the training data on a multinomial logistic regression model. This would make accurate predictions on unseen data feasible. One Bayesian approach for this problem is to use a prior distribution for $\beta_j$ which assigns a high probability to values of $\beta_j$ near $\mu_j$. Since each individual covariate cannot be expected to exert much leverage on the response, one may choose $\mu_{ji}$ in the vicinity of zero and $\sigma_{ji}$ a small number, at most 2 to reflect the prior belief on $\beta_{ji}$. These prior beliefs can be reflected by a normal prior $\beta_j \sim N[\boldsymbol{\mu}_j, \Sigma_j]$ which has been frequently used in the literature. Therefore, the choice of a set of values for $\mu_j$ and $\Sigma_j$, e.g.$(\mathbf{0}, I)$, can help initiate the computational process described below. If someone does not prefer to use this normal prior due to lack of prior information, he may choose a non-informative prior like uniform in place of normal. In that case, an alternative estimator would be obtained while the rest of our procedure for estimating $\hat{P}_j$ would be the same as for the normal prior. We assume a priori that the components of $\beta_j$ are independent and hence the overall prior for $\beta_j$ is the product of the priors for its components. Hence, in addition to the likelihood function (2.3), and the assumed structure for the logistic regression in (2.2), the prior distributions are specified below:

$$\beta_j \sim N[\boldsymbol{\mu}_j, \Sigma_j], \qquad\qquad j = 1, 2, \cdots, K \qquad\qquad (2.4)$$

where $\Sigma_j = Diag(\sigma_{j0}, \cdots, \sigma_{jq})$. Thus the posterior distribution is obtained as

$$\pi(\beta|\mathbf{Y} = \mathbf{y}) \propto \prod_{i=1}\left\{\frac{exp\left(\sum_{j=1}(y_{ij}\mathbf{x}_i'\beta_j - \frac{1}{2}((\beta_j - \mu_j)'\Sigma_j^{-1}(\beta_j - \mu_j)))\right)}{(1 + \sum_{j=1} exp(\mathbf{x}_i'\beta_j))^{N_i}}\right\}$$
$$(2.5)$$

Bayesian inference is now based on the analysis of the posterior distribution (2.5). In general, this posterior will not have a known closed form rather it will have a complicated high dimensional density only known up to the normalizing constant which makes direct inferences almost impossible. Markov Chain Monte Carlo(MCMC) methods are techniques that have been developed to resolve this kind of problem.

Thus, we employ the MCMC method to compute the posterior distribution and the posterior expectations of some functions of $\beta_j$.

We have to compute

$$m(\mathbf{y}) = \prod_i \frac{N_i!}{\prod_j y_{ij}!} \times$$

$$\prod_i \int_{\boldsymbol{\beta}_1} \cdots \int_{\boldsymbol{\beta}_K} \frac{exp\Big(\sum_{j=1}(y_{ij}\mathbf{x}_i'\boldsymbol{\beta}_j - \frac{1}{2}((\boldsymbol{\beta}_j - \boldsymbol{\mu}_j)'\Sigma_j^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j)))\Big)}{(1 + \sum_{j=1} exp(\mathbf{x}_i'\boldsymbol{\beta}_j))^{N_i}} d\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_K$$

(2.6)

as the marginal distribution of **y** and

$$\pi(\beta|Y = y) = \frac{1}{m(\mathbf{y})}$$

$$\prod_i \Big(\frac{N_i!}{\prod_j y_{ij}!}\Big) \frac{exp\Big(\sum_{j=1}(y_{ij}\mathbf{x}_i'\beta_j - \frac{1}{2}((\beta_j - \mu_j)'\Sigma_j^{-1}(\beta_j - \mu_j)))\Big)}{(1 + \sum_{j=1} exp(\mathbf{x}_i'\beta_j))^{N_i}}. \quad (2.7)$$

as the posterior density of $\beta_j$. Now, computation of Bayes estimator is quite simple. The Bayes estimator of $\beta_j$ is any number $\hat{\beta}_j$, minimizing the posterior risk $E(L(\beta_j, \hat{\beta}_j)|\mathbf{Y} = \mathbf{y})$. If we define,

$$L(\beta_j, \hat{\beta}_j) = \begin{cases} 0, & |\beta_j - \hat{\beta}_j| < C \\ \\ 1, & |\beta_j - \hat{\beta}_j| \geq C, \end{cases}$$

then, $\hat{\beta}_j$ is the midpoint of the interval $I = \{\hat{\beta}_j, |\beta_j - \hat{\beta}_j| \leq C\}$ of length $2C$ which maximizes $P(\hat{\beta}_j \in I|y)$, (Robert, 2001, pp. 166-167). In this case the Bayes estimator is the posterior mode. From (2.7) we have

$$log\Big(\pi(\beta|Y = y)\Big) = A + \sum_{i=1}\sum_{j=1}\Big(y_{ij}\mathbf{x}_i'\beta_j - \frac{1}{2}((\beta_j - \mu_j)'\Sigma_j^{-1}(\beta_j - \mu_j))\Big)$$

$$- \sum_{i=1} N_i log(1 + \sum_{j=1} exp(\mathbf{x}_i'\beta_j)) \qquad (2.8)$$

where

$$A = log\Big(\frac{1}{m(\mathbf{y})}\prod_i\Big(\frac{N_i!}{\prod_j y_{ij}!}\Big)\Big).$$

Finding the maximum of (2.7) or equivalently of (2.8), gives the Bayes estimator of $\beta_j$, as

$$\hat{\beta}_j = \mu_j + \frac{1}{I}\Sigma_j\Big(\sum_{i=1} a_{ij}\mathbf{x_i}\Big) = \mu_j + \frac{1}{I}\Sigma_j\mathbf{X}'\Big(\mathbf{Y}_j - N_i\mathbf{P}_j\Big). \qquad (2.9)$$

where $a_{ij} = y_{ij} - N_i p_{ij}$.

To appreciate the nature of the estimator of $\beta_j$, we present it in an iterative format. For any given $\mathbf{P}_j$, use an initial estimator $\hat{\mathbf{P}}_j$, satisfying $n^{\frac{1}{2}}(\hat{\mathbf{P}}_j - \mathbf{P}_j) = O_p(1)$ to obtain an estimator $\hat{\beta}_j$ of $\beta_j$. Such initial estimators can easily be obtained. Our method starts with the likelihood estimator and iterates the following steps 1-4 until it converges:

**Step 1:** Let $\hat{\mathbf{P}}_j^{(0)} = \frac{1}{N_i}\mathbf{Y_j}$

**Step 2:** Given $\hat{\mathbf{P}}_j^{(0)}$, let $\hat{\beta}_j^{(0)}$ be the solution of (2.9) for $\beta_j$.

**Step 3:** Find $\hat{\boldsymbol{\eta}}_j = X\hat{\beta}$ from (2.2).

**Step 4:** Given $\hat{\boldsymbol{\eta}}_j$, let the updated estimator of $\hat{\mathbf{P}}_j$ and $\hat{\beta}_j$ be $\hat{\mathbf{P}}_j^{(1)}$ and $\hat{\beta}_j^{(1)}$.

Iterate steps 2-4 until a convergence criterion is met.

In case of non-availability of good prior information to formulate an informative prior, we can employ a non-informative prior for $\beta_j$ such as uniform prior for each $\beta_j$. The resulting posterior density is then proportional to the likelihood function:

$$\pi(\beta|\mathbf{Y} = \mathbf{y}) \propto \prod_{i=1} L(\beta|\mathbf{Y}_i = \mathbf{y}_i)$$

Hence, the maximum likelihood estimator of $\beta_j$, is the Bayesian estimator under the 0-1 loss function. The Newton-Raphson method is a suitable iterative method for solving this problem. Thus, we can write the following algorithm:

**Step 1:** Let $\hat{\mathbf{P}}_j^{(0)} = \frac{1}{N_i}\mathbf{Y}_j$
For $t = 0, 1, \cdots, T$ we have,

**Step 2:** $\mathbf{Z}_j^{(t)} = \hat{\boldsymbol{\eta}}_j^{(t)} + Diag^{-1}(N_i\hat{P}_{ij}^{(t)}(1 - \hat{P}_{ij}^{(t)}))(\mathbf{Y}_j - \mathbf{N}_i\hat{\mathbf{P}}_j^{(t)})$

**Step 3:**

$$\hat{\beta}_j^{(t+1)} = (\mathbf{X}'Diag(N_i\hat{P}_{ij}^{(t)}(1-\hat{P}_{ij}^{(t)}))\mathbf{X})^{-1}(\mathbf{X}'Diag(N_i\hat{P}_{ij}^{(t)}(1-\hat{P}_{ij}^{(t)}))\mathbf{Z}_j^{(t)})$$

$$(2.10)$$

**Step 4:** $\hat{P}_{ij}^{(t+1)} = \frac{exp(\mathbf{x}_i'\hat{\beta}_j^{(t+1)})}{1+exp(\mathbf{x}_i'\hat{\beta}_j^{(t+1)})}$.

To provide an idea about the asymptotic behavior of the posterior distribution, one notes that when $N_i \to \infty$, each $Mult(N_i, P_i)$ tends to a multivariate normal distribution. Since the chosen prior is either normal or a constant the asymptotic posterior will be a normal. This fact has generally been proven for posterior distributions, Bernardo and smith (1994).

In order to estimate $\beta_j$, we employ the MCMC method. The MCMC techniques generate samples from desired distributions by embedding them as limiting distributions of a Markov chain. There are many ways of categorizing MCMC methods. The simplest one is to classify them in one of two groups: the first group is used in estimation problems where the unknowns are typically parameters of a model, which are assumed to have generated the observed data; the second group is employed in more general scenarios where the unknowns are not only model parameters, but the models as well. In this paper, we use the MCMC method for both groups. On the other hand, we construct Markov chains that have as their stationary distribution the required posterior distribution of the parameters. Details of the technicalities involved for MCMC can be found, for example, in Rosenberg and Young(1999) and Ntzoufras et al.(1999). One of the most popular MCMC methods, is the Gibbs sampler. Here, we will estimate the marginal and the posterior distributions of parameters using Gibbs sampling.

First, we estimate the marginal distribution. The marginal distribution can be computed from the $n$ realizations of the Gibbs sequence. For $j = 1, \cdots, K$, if we draw a large number of values $\beta_{1j}, \beta_{2j}, \cdots, \beta_{nj}$ from the density (2.4), then from (2.6) we shall have:

$$m(y) = E_{\pi(\boldsymbol{\beta})}\Big( L((\beta_1, \cdots, \beta_K)|Y = y) \Big) \simeq \frac{1}{n} \sum_{i=1}^{n} L((\beta_{i.j}, \cdot, \beta_{i.K})|Y = y).$$

(2.11)

The estimator (2.11) is unstable when the prior is diffuse or the likelihood is much more concentrated than the prior. In such cases the simulation will be inefficient since most of the simulated values will have low likelihood values and therefore the estimator will be dominated by few large values. Moreover, the variance of the estimator (2.11) will be large and the convergence of the estimator to its true value will be very slow. An alternative way to approximate the marginal distribution is the Laplace approximation. This method has been used by Tierney and Kadane (1986), Erkanli(1994) and Kadane

and Lazar(2004) for a model of dimension $d$. Its application provides

$$log\Big(m(\mathbf{y})\Big) \simeq 0.5\, d\, log(2\pi) + 0.5 log|H^*| + log\Big(L(\beta^*|Y = y)\pi(\beta^*)\Big)$$
(2.12)

where $\beta^*$ is the vector of posterior mode estimate of $\beta$, (Lewis and Raftery, 1997) and $H^*$ is the inverse of the Hessian matrix $\{\partial^2 h(\beta, \mathbf{y})/ \partial\beta\, \partial\beta'\}$ of $h(\beta, \mathbf{y}) = log\Big(L(\beta|Y = y)\pi(\beta)\Big)$ evaluated at $\beta^*$.

Usually, the Bayes factor is used for models comparison. If we apply the above approximation by expanding the numerator and denominator of the Bayes factor, we would get an approximation of the Bayes factor. For the models $M_i$ and $M_{i'}$, one has

$$BF_{ii'} \simeq \frac{L(\beta_i^*|Y = y)}{L(\beta_{i'}^*|Y = y)}\left|\frac{H^*(\beta_i^*)}{H^*(\beta_{i'}^*)}\right|^{1/2}\left(\frac{1}{2\pi}\right)^{(d_{i'} - d_i)/2}$$
(2.13)

or

$$log(BF_{ii'}) \simeq log(\lambda_n) + S(\beta_i^*, \beta_{i'}^*)$$
(2.14)

where $\lambda_n$ is the standard likelihood ratio for the comparison of models $M_i$ and $M_{i'}$, and $S(\beta_i^*, \beta_{i'}^*)$ denotes the remainder term. When $M_i$ is a submodel of $M_{i'}$, the remainder term $S(\beta_i^*, \beta_{i'}^*)$ is of $o(1)$, (see Gelfand and Dey (1994)).

To compare various models by the Bayes factor, we use the Bayes factor approximation (2.13). Now, (2.13) along with its interpretations provided by Kass and Raftery (1995), is used to choose the best model for an specific example, i.e., the pregnancy outcomes in consanguineous marriages.

# 3 Analysis of Pregnancy Outcomes in Consanguineous Marriages

In this Section, we reanalyze a subset of data regarding the pregnancy outcome in consanguineous marriages. Schull(1958) analyzed these data using a frequentist approach and Ntzoufras et al.(1999) reanalyzed it via Metropolis-Hastings algorithms. We will analyze it by a Bayesian procedure outlined above. Data in Table 1, taken from Ntzoufras et al.(1999), which initially appeared in Schull(1958), show the pregnancy outcomes in various districts for different degrees of consanguinity. This sample, according to the degree of consanguinity between the parents, included 6358 pregnant women in

three districts of Shizuka city, Japan. Here, we have two covariates, R≡Residence( coded as Rural district≡-1, Intermediate district≡0, Urban district≡1) and C≡Consanguinity( coded as no relation≡0, 2nd cousins≡1, $1\frac{1}{2}$ cousins≡1.5, 1st cousins≡2). The categories of the response variable are Deaths( A≡Abortion, S≡ Stillbirth, U≡ death in less than 12 months, V≡ death in 13-60 months, and Su≡ Survived). We consider a multinomial regression model (2.2) for the pregnancy outcome, with district as a categorical covariate. The goal is to find the effect of consanguinity on the pregnancy outcome while accounting for the type of residence district. Here, we shall use the *Survived* as the baseline category in the model (2.2) which is now stipulated as

$$M_1: \qquad \eta_{ij} = log\left(\frac{p_{ij}}{p_{i,su}}\right) = \beta_{j0} + \beta_{j1}R + \beta_{j2}C + \beta_{j3}R*C \quad (3.1)$$

for $j = A, S, U, V$. Also we consider three submodels as alternatives to $M_1$:

$$M_2: \qquad \eta_{ij} = \beta_{j0} + \beta_{j1}R + \beta_{j2}C \qquad\qquad (3.2)$$

$$M_3: \qquad \eta_{ij} = \beta_{j0} + \beta_{j1}R \qquad\qquad\qquad (3.3)$$

$$M_4: \qquad \eta_{ij} = \beta_{j0} + \beta_{j2}C \qquad\qquad\qquad (3.4)$$

Now, we compute the log Bayes factors for these models by (2.14) with respect to $M_2$ for various categories of death, which are presented in Table 2. It can be observed that, $M_2$, is the best one for data under study. For this model, we present the results of Bayesian Markov Chain Monte Carlo analysis. The procedure employed here follows the method of Bayesian logistics regression analysis. From (2.9) and using $exp(\hat{\eta}_{ij} - \hat{\eta}_{ik}) = (\frac{\hat{p}_{ij}}{\hat{p}_{ik}})$, $i \neq k = 1, \cdots, K$, we have obtained the estimates of the coefficients in equation (3.2) and confidence region , respectively. A confidence region is the highest posterior density (HPD) region, that is,

$$B = \left\{\beta; \pi(\beta|y) > c\right\}$$

where $B$ has a posterior probability $(1 - \alpha)$ of containing $\beta$. These values are shown in Table 3. Having obtained $\hat{\beta}_j = (\hat{\beta}_{j0}, \hat{\beta}_{j1}, \hat{\beta}_{j2}, \hat{\beta}_{j3})$, we can estimate $\hat{\eta}_{ij}$, from (2.2). Next, we compute various log odds relative to various levels of covariates from (2.2). These log odds are given in Tables 4-7. Finally, Table 8 provides the estimated odds ratios with respect to the baseline category, "no relation". We see that Consanguineous marriages and Residence independently have significant effect on the pregnancy outcome for each cell.

**Table 1.** Pregnancy Outcomes in Consanguinous Marriages(Data from Ntzoufras et al.(1999))

| Resid. | Consang. | Abor | Still | $\leq 12Mon.$ | $13-60Mon.$ | Survived |
|--------|----------|------|-------|---------------|-------------|----------|
| | no relation | 27 | 15 | 57 | 25 | 834 |
| Rural | 2nd cousins | 1 | 1 | 13 | 6 | 139 |
| district | $1\frac{1}{2}$ cousins | 3 | 2 | 7 | 2 | 51 |
| | 1st cousins | 12 | 2 | 18 | 11 | 250 |
| | no relation | 67 | 20 | 128 | 76 | 2379 |
| Inter. | 2nd cousins | 11 | 1 | 25 | 10 | 291 |
| district | $1\frac{1}{2}$ cousins | 11 | 4 | 14 | 12 | 196 |
| | 1st cousins | 23 | 6 | 40 | 27 | 558 |
| | no relation | 7 | 5 | 21 | 14 | 496 |
| Urban | 2nd cousins | 4 | 0 | 1 | 2 | 63 |
| district | $1\frac{1}{2}$ cousins | 3 | 0 | 5 | 2 | 100 |
| | 1st cousins | 7 | 1 | 15 | 11 | 226 |

**Table 2.** Estimates of the log Bayes factor approximations from (2.13).

| outcome. | $M_1$ | $M_2$ | $M_3$ | $M_4$ |
|----------|-------|-------|-------|-------|
| Abortion | .211 | 1 | .040 | .010 |
| Stillbirth | .453 | 1 | .384 | .341 |
| $\leq 12Mon.$ | .329 | 1 | .400 | .412 |
| $13-60Mon.$ | .112 | 1 | .165 | .185 |

For this problem, the design matrix is

$$X = \left( \begin{array}{ccc} \mathbf{1} & -\mathbf{1} & \mathbf{M} \\ \mathbf{1} & \mathbf{0} & \mathbf{N} \\ \mathbf{1} & \mathbf{1} & \mathbf{P} \end{array} \right)$$

with

$$\mathbf{1} = \left( \begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array} \right), \quad \mathbf{M} = \left( \begin{array}{cc} 0 & 0 \\ 1 & -1 \\ 1.5 & -1.5 \\ 2 & -2 \end{array} \right),$$

$$\mathbf{N} = \left( \begin{array}{cc} 0 & 0 \\ 1 & 0 \\ 1.5 & 0 \\ 2 & 0 \end{array} \right), \quad \mathbf{P} = \left( \begin{array}{cc} 0 & 0 \\ 1 & -1 \\ 1.5 & 1.5 \\ 2 & 2 \end{array} \right)$$

**Table 3.** The Bayesian MCMC estimated parameters of the log odds model and their confidence region(baseline category="survived")

| Outcome | Source | coefficient | conf. region(HPD) |
|---------|--------|-------------|-------------------|
| Abortion | Cons. | -3.037 | (-4.124 -2.143) |
| | R | -0.438 | (-0.534 -0.343) |
| | C | 0.394 | (0.154 0.432) |
| | R*C | -0.092 | (-0.524 0.214) |
| Stillbirth | Cons. | -4.828 | (-5.765 -3.789) |
| | R | -0.592 | (-0.625 -4.987) |
| | C | 0.81 | (0.787 0.943) |
| | R*C | 0.001 | (-0.001 0.002) |
| $\leq 12$Months | Cons. | -2.631 | (-3.546 -1.943) |
| | R | 0.08 | (0.076 0.091) |
| | C | -0.15 | (-0.173 -0.125) |
| | R*C | -0.197 | (-0.256 0.129) |
| 13-60 months | Cons. | -3.124 | (-4.765 2.87) |
| | R | -0.101 | (-0.241 -0.043) |
| | C | -0.384 | (-0.424 -0.245) |
| | R*C | -0.061 | (-0.283 0.098) |

**Table 4.** The Bayesian MCMC log odds of "abortion" to "survived"

| *Residence* | *consanguinity* | | | |
|---------|---------------|----------------|---------------------|--------------|
| | *no relation* | *2nd cousins* | *$1\frac{1}{2}$ cousins* | *1st cousins* |
| *Rural* | $-3.981$ | $-3.803$ | $-3.333$ | $-3.521$ |
| *Intermediate* | $-3.574$ | $-3.398$ | $-2.921$ | $-3.132$ |
| *Urban* | $-3.551$ | $-3.351$ | $-2.897$ | $-3.097$ |

**Table 5.** The BMCMC log odds of "stillbirth" to "survived"

| *Residence* | *consanguinity* | | | |
|---------|---------------|----------------|---------------------|--------------|
| | *no relation* | *2nd cousins* | *$1\frac{1}{2}$ cousins* | *1st cousins* |
| *Rural* | $-5.021$ | $-6.102$ | $-4.466$ | $-5.176$ |
| *Intermediate* | $-4.711$ | $-5.667$ | $-4.089$ | $-4.854$ |
| *Urban* | $-4.066$ | $-5.144$ | $-3.535$ | $-4.251$ |

**Table 6.** The BMCMC log odds of "$\leq 12$ months" to "survived"

| Residence | consanguinity | | | |
|---|---|---|---|---|
| | no relation | 2nd cousins | $1\frac{1}{2}$ cousins | 1st cousins |
| Rural | $-3.185$ | $-2.854$ | $-2.838$ | $-2.945$ |
| Intermediate | $-2.898$ | $-2.564$ | $-2.552$ | $-2.651$ |
| Urban | $-2.711$ | $-2.369$ | $-2.361$ | $-2.452$ |

**Table 7.** The BMCMC log odds of "13-60 months" to "survived"

| Residence | consanguinity | | | |
|---|---|---|---|---|
| | no relation | 2nd cousins | $1\frac{1}{2}$ cousins | 1st cousins |
| Rural | $-3.596$ | $-3.433$ | $-3.181$ | $-3.174$ |
| Intermediate | $-3.422$ | $-3.261$ | $-3.125$ | $-3.161$ |
| Urban | $-3.617$ | $-3.222$ | $-3.119$ | $-3.080$ |

**Table 8.** The Bayesian MCMC estimated log odds with respect to "no relation" for each district

| Consan. | Resid. | Rural | Interm. | Urban |
|---|---|---|---|---|
| Abortion | 2nd cousins | .920 | .840 | .810 |
| | 1.5st cousins | .523 | .522 | .518 |
| | 1st cousins | .633 | .648 | .635 |
| Stillbirth | 2nd cousins | .339 | .384 | .341 |
| | 1.5st cousins | 1.748 | 1.860 | 1.715 |
| | 1st cousins | .860 | .911 | .833 |
| $\leq 12Mon.$ | 2nd cousins | 1.394 | 1.400 | 1.412 |
| | 1.5st cousins | 1.408 | 1.415 | 1.423 |
| | 1st cousins | 1.254 | 1.277 | 1.295 |
| $13 - 60Mon.$ | 2nd cousins | 1.181 | 1.165 | 1.185 |
| | 1.5st cousins | 1.511 | 1.500 | 1.522 |
| | 1st cousins | 1.528 | 1.545 | 1.564 |

Now, based on developed theory, the models in Table 3 are studied numerically. Considering the stillbirth case as an example, the obtained model will be as follows:

$$log\left(\frac{\hat{P}_s}{\hat{P}_{Su}}\right) = -4.828 - 0.592R + 0.81C$$

This model is considered in two different cases:

**Case 1:** Keeping consanguinity fixed at one of its levels, we obtained the odds for various level of R. For example, the odds of stillbirth in rural to urban district is:

$$OR(S, Su|C) = \frac{exp(-4.828 - 0.592(-1) + 0.81C)}{exp(-4.828 - 0.592(1) + 0.81C)} = 3.267$$

which shows the rural residents have 3.267 times more chances of stillbirth than urban residents, given they have the same degree of consanguinity.

**Case 2:** Keeping the type of residence fixed ,one can compute the odds of stillbirth for first cousins as compared to no relation:

$$OR(S, Su|R) = \frac{exp(-4.828 - 0.592R + 0.81(1))}{exp(-4.828 - 0.592R + 0.81(0))} = 2.247$$

this number shows that among the residents of a given place, the first cousin-parents have 2.247 times more probability to have stillbirth compared to no-related parents. In general accounting for the effect of residential place, one can conclude that the consanguinity is an effective factor in child mortality.

## 4    A Small Scale Simulation

As suggested by a referee,we perform a small scale simulation to evaluate the performance of our proposed algorithm. In this simulation we choose

$$\mathbf{P} = \left( \begin{array}{ccc} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.2 & 0.4 \\ 0.7 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{array} \right).$$

To generate the data according to the multinomial distribution with cell probabilities given in various rows of $\mathbf{P}$, we sampled 40,50,50 and 50 observations from $U(0,1)$ and assigned them to rows of $Y$.

$$\mathbf{Y} = \left( \begin{array}{ccc} 8 & 17 & 15 \\ 19 & 14 & 17 \\ 36 & 5 & 9 \\ 17 & 16 & 17 \end{array} \right).$$

suppose there two binary covariates. The proposal model contains main effects and the interaction term. Thus according to(2.2), the design matrix is

$$\mathbf{X} = \left( \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array} \right).$$

Applying our algorithm leading to (2.10), after 10 iterations, we obtained

$$\hat{\beta}_0 = \left( \begin{array}{c} -0.6007739 \\ 1.9467943 \\ 0.7089874 \\ -2.0550079 \end{array} \right).$$

$$\hat{\beta}_1 = \left( \begin{array}{c} 0.1212609 \\ -0.6679046 \\ -0.3094131 \\ 0.7971163 \end{array} \right).$$

Even for this small number of iterations, our algorithm has lead to quite acceptable results for reconstructing $\mathbf{P}$ as $\mathbf{P}^B$

$$\mathbf{P}^B = \left( \begin{array}{ccc} 0.2098 & 0.3822 & 0.4080 \\ 0.3861 & 0.2871 & 0.3268 \\ 0.7221 & 0.1089 & 0.1691 \\ 0.3465 & 0.3267 & 0.3268 \end{array} \right).$$

Using $\mathbf{P}^B$, we have computed the cell expected values, which are used for testing the goodness of fit of our model.

$$\mathbf{E}^B = \left( \begin{array}{ccc} 8.395 & 15.288 & 16.321 \\ 19.307 & 14.355 & 16.340 \\ 36.139 & 5.445 & 8.45 \\ 17.327 & 16.337 & 16.340 \end{array} \right).$$

$\chi^2 = 0.89272$. As a final observation, one can see the proposed algorithm performs well in estimating $\mathbf{P}$ and the expected cell counts.

# References

Bernardo, J. M. and Smith, A. F. M. (1994), Bayesian Theory. First edition, Chichester: Wiley.

Chen, M. H. and Dey, D. K. (1998), Bayesian modelling of correlated binary responses via scale mixture of multivariate normal link function. Sankya: The Indian Journal of Statistics, **60**, 322-343.

Dellaportas, P. and smith, A. (1993), Bayesian inference for genaralized linear and proportional hazard models via Gibbs sampling. Appl. Statist., **42**, 443-459.

Erkanli, A. (1994), Laplace approximations for posterior expectation when the model occurs at the boundary of the parameter space. Journal of the American Statistical Association, **89**, 250-258.

Eskandari, F. and Meshkani, M. R. (2004), Empirical Bayes analysis of log-linear models for generalized finite stationary Markov chain. Metrika, **59**, 173-191.

Gamerman, D. (1997), Efficient sampling from the posterior distribution in generalized linear mixed models. Statitics and Computing, **7**, 57-68.

Gelfand, A. E., Smith, A. F. M., and Lee T. M. (1992), Bayesian analysis of constrained parameter and trancleted data problems using Gibbs sampling. Journal of the American Statistical Association, **87**, 523-532.

Gelfand, A. E. and Dey, D. K. (1994), Baysian model choice: asymptotics and exact calculations. J. Roy. Statist. Soc. Ser. B, **56**, 501-514.

Gelfand, A. E and Ghosh, S. K. (1998), Model choice: a minimum posterior predictive loss approach. Biometrika, **85**, 1-11.

Kadane, J. and Lazar, N. (2004), Methods and criteria for model selection. Journal of the American Statistical Association, **99**, 279-290.

Kass, R. E. and Raftery, A. E. (1995), Bayes factors. Journal of the American Statistical Association, **90**, 773-795.

Laud, P. W. and Ibrahim, J. G. (1995), Predictive model selection. Journal of the Royal Statistical Socity. B, **57**, 247-262.

Lewis, S. and Raftery, A. (1997), Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. Journal of the American Statistical Association, **92**.

Madigan, D., Genkin, A., Lewis, D. D., Fradkin, D. (2006), Bayesian Multivariate Logistic Regression for Author Identification. Technical Report. Department of Stat., Rutgers University.

Ntzoufras, I., Dellaportas, P., and Forster, J. J. (1999), Specification and interpretation of prior distributions for variable selection in linear model. Proceedings of the Fourth Helenic-European Conference on Computer Mathematics and its Applications, (E. A. Lipitakis, ed.).

Robert, C. P. (2001), The Bayesian Choice. 2nd edition, New York: Springer.

Rosenberg M. A. and Young V. R.(1999), A Bayesian approach to understanding time series data. North American Actuarial Journal, **3**(2), 130-143.

Schmidt, P. and Strauss, R. (1975), The prediction of occupation using multiple logit models. Internat. Econ Rev., **16**, 471-486.

Schull, W. J. (1958), Empirical risks in consanguineous marriages: sex ratio, malformation and viability. Amer. Jour. of Human Genetics, **30**, 294-212.

Tierney, L. and Kadane, J. (1986), Accurate Approximations for posterior Moments and Marginal Densities. Journal of the American Statistical Association, **81**, 82-86.

Wang, k. and Kapat, p. (2005), Classification Using Bayesian Logistic Regression. Technical Report, Department of Stat., Ohio Univrsity.