Journal of the Korean
Statistical Society
Vol. 26, No. 4, 1997

Sampling Based Approach to Bayesian Analysis of Binary Regression Model with Incomplete Data †

Younshik Chung[1]

Abstract

The analysis of binary data appears to many areas such as statistics, biometrics and econometrics. In many cases, data are often collected in which some observations are incomplete. Assume that the missing covariates are missing at random and the responses are completely observed. A method to Bayesian analysis of the binary regression model with incomplete data is presented. In particular, the desired marginal posterior moments of regression parameter are obtained using Meterpolis algorithm (Metropolis et al. 1953) within Gibbs sampler (Gelfand and Smith, 1990). Also, we compare logit model with probit model using Bayes factor which is approximated by importance sampling method. One example is presented.

Key Words : Binary regression model; Logit model; Probit model; Missing at ramdom; Metropolis algorithm; Gibbs sampler; Importance sampling method; Bayes factor.

---

†This reserach was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1995.

[1] Department of Statistics, Pusan National University, Pusan, 609-735 Korea.

1. INTRODUCTION

Generalized linear model (GLM : McCullagh and Nelder, 1989) is an extension of the classical linear models and so have unified regression methodology. Regression and discrimination using probit and logit models have become increasingly popular with the easy availability of appropriate computer routines. The classical approach fits a binary response regression model using maximum likelihood, and inferences about the model are based on the associated asymptotic theory. In the statistical literature on the analysis of independent data, EM algorithm (Dempster, Laird and Rubin, 1977) was employed to many areas. For examples, Fuchs (1982) for problems of incomplete data in log-linear models, and Ibrahim (1990) for incomplete data in GLM. Also, Vach and Schmacher (1993) considered logistic regression with incompletely observed categorical covariates. Our focus is a fully Bayesian parametric approach. Use of the Bayesian framework for inference with regard to the GLM only with complete data dates back to work by Stephens and Dellaportas (1992), Albert and Chib (1993), Dellaportas and Smith (1993). In our view, wider use of the Bayesian framework has been impeded by the difficulties in computing required marginal posterior distributions of the model parameters. Recently developed methods in computation techniques, so called the Gibbs sampler, an iterative Monte Carlo method, avoids sophisticated analytic and numerical high dimensional integration procedures. The conceptual simplicity of Gibbs sampler may prove an attractive alternative to the analytic and/or numerical sophistication demanded by other methods.

In particular, in section 2 we formulate the Bayesian binary regression model with incomplete data. We clarify what distributions are sought and what distributions can be readily sampled. Also the estimate of marginal posterior density is proposed without using kernel density estimation. In section 3, Bayesian model selection is explained using the approximating Bayes factor. In section 4, we present one example.

2. BAYESIAN FORMULATION FOR BINARY REGRESSION MODEL

Densities are denoted generically by brackets, so joint, conditional, and marginal forms, for example, appear as $[X, Y]$, $[X|Y]$ and $[X]$, respectively. Suppose that $y_1, \ldots, y_n$ are independent observations having a density in the exponential family of the form

$$[y_i|\theta_i,\phi] = \exp\{(y_i\theta_i - b(\theta_i))/a_i(\phi) + c(y_i,\phi)\} \tag{2.1}$$

for some functions $a_i(\cdot)$, $b(\cdot)$ and $c(\cdot)$. The parameters θ_i is called the canonical parameter for the ith observation and ϕ is a scalar dispersion parameter distinct from θ_i. Now consider the GLM with $g(\mu_i) = \eta_i = x_i^t\beta$ where x_i^t is a $1 \times p$ vector corresponding to the ith row of $n \times p$ matrix of covariates X, β is a $p \times 1$ vector of regression parameter and g is a monotonic differential function(the link function), and $\mu_i = E(Y_i)$. In GLM, if the relationship between θ_i and β is of the form $\theta_i = h(x_i^t\beta)$ for some function h depending on the link function g, the joint probability density of n independent observations $Y = (y_1,\ldots,y_n)$ can be written as

$$[Y|X,\beta] = \prod_{i=1}^{n} exp\{(y_i h(x_i^t\beta) - b(h(x_i^t\beta)))/a_i(\phi) + c(x_i,\phi)\}. \tag{2.2}$$

For the binary regression model, suppose that n independent binary random variables $Y_1,\ldots,Y_n$ are observed, where Y_i is distributed Bernoulli with probability of success p_i. Then the p_i are related to a set of covariates that may be continuous or discrete. Define the binary regression model as $p_i = H(x_i^t\beta)$, $i = 1,\ldots,n$ where β is a $p \times 1$ vector of unknown parameters, x_i^t is a vector of known covariates and H is a known cdf linking the probabilities p_i with the linear structure $x_i^t\beta$.

The probit model is obtained if H is the standard Gaussian cdf, while the logit model is obtained if H is the logistic cdf. We consider a binary regression model with m types of covariates and l missing data among n total covariate data where m is the number of all possible combinations of covariates. Without loss of generality, reorder the data so that the first $n - l$ data are completely observed and the remaining l data are unobserved (incomplete).

The essential idea is quite simple. Suppose that along with the missing covariate x_i, we have the corresponding latent data, say z_i (see Tanner and Wong, 1987). So, we assume that the missing covariates are missing at random(MAR) and the responses(y's) are completely observed. Also we assume that all covariates are dichotomized. So we specify a joint distribution on X_i as $[x_i|\gamma] = \prod_{j=1}^{m} \gamma_j^{I_j(x_i)}$ where $I_j(x_i)$ is an indicator function for the jth sequence of covariate vector $x_i^t (j = 1,\ldots,m)$ and $\sum_{j=1}^{m} \gamma_j = 1$.

Let $[\beta|\delta]$ be a proper or improper density which summarizes our prior information about β with hyperparameter δ. At the third stage, we may consider the distribution $[\delta]$ of δ. In this situation, if our interest centers on the marginal distribution of β_i, the integration operations necessary are not analytically tractable, and so we are forced to use computational techniques. Then the joint density for Y, β, δ and γ is of the form

$$\begin{aligned}
[Y,\beta,\delta,\gamma|X] &= [Y|X,\beta][\beta|\delta][\delta][\gamma] \\
&\propto \prod_{i=1}^{n} exp\{(y_i h(x_i^t\beta) - b(h(x_i^t\beta)))/\phi + c(x_i,\phi)\} \\
&* \prod_{i=1}^{l}\prod_{j=1}^{m} \gamma_j^{I_j(z_i)} * [\beta|\delta] * [\delta] * [\gamma]. \qquad (2.3)
\end{aligned}$$

The marginal posterior of β_i is given by

$$\begin{aligned}
[\beta_i] = [\beta_i|Y,X] &= \int [\beta|Y,X] \textstyle\prod_{j\neq i} d\beta_j \\
&= \int \frac{\int [Y,\beta,\delta,\gamma|X] d\gamma d\delta}{\int\int [Y,\beta,\delta,\gamma|X] d\beta d\gamma d\delta} \textstyle\prod_{j\neq i} d\beta_j. \qquad (2.4)
\end{aligned}$$

The joint posterior distribution of β given the data is proportional to expression (2.3). However, marginalization of expression (2.3) requires integration over δ and γ which is generally a formidable analytic problem. Use of Gibbs sampler, as discussed in the context of hierarchical Bayesian models in Gelfand and Smith (1990) enables a straightforward sampling-based solution to such problems. Implementation requires sampling from the complete conditional distributions $[\gamma|Y,X,\beta]$ and $[\beta_i|Y,X,\gamma,\beta_j, j\neq i]$, $i = 1,\ldots,k$. Each of these distributions is also proportional to expression (2.4). Random generation methods such as Metropolis algorithm (Metropolis et al. 1953) enable such sampling. Since the distribution of y_i given x_i and β is Bernoulli with probability of success p_i, assume that for logit or probit model, $p_i = exp(x_i{}^t\beta)(1+exp(x_i{}^t\beta))^{-1}$ or $p_i = \Phi(x_i{}^t\beta)$, respectively where $x_i{}^t = (1, x_{i1}, \ldots, x_{ip-1})$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^t$ and Φ denotes the cumulative distribution function of standard normal variable. Assume that β is distributed to the normal density with mean vector $\mu = (\mu_0, \mu_1, \ldots, \mu_{p-1})$ and the covariance-variance matrix Σ and $[\gamma]$ is Dirichlet. That is,

$$[\gamma|\alpha_1,\ldots,\alpha_{m+1}] \propto \gamma_1^{\alpha_1-1}\cdots\gamma_m^{\alpha_m-1}(1-\sum_{i=1}^{m}\gamma_i)^{\alpha_{m+1}-1}. \qquad (2.5)$$

For $i = 1,\ldots,l$, let $Z_i = X_{n-l+i}$. That is, Z_i's denote the covariates with missing data and then $X = (X_1,\ldots,X_{n-l}, Z_1,\ldots,Z_l)$.

2.1 Logit model

Since $p_i = exp(x_i{}^t\beta)(1+exp(x_i{}^t\beta))^{-1}$, the joint density for Y, β, δ and γ is of the form

$$\begin{aligned}[Y,\beta,\delta,\gamma|X] &= [Y|X,\beta][\beta|\delta][\delta][\gamma] \\ &\propto exp\left[\sum_{j=1}^{n} y_i(\beta_0+\sum_{i=1}^{p-1} x_{ij}\beta_i) - \sum_{j=1}^{n} log(1+exp(\beta_0+\sum_{i=1}^{p-1} x_{ij}\beta_i))\right] \\ &* \prod_{i=1}^{l}\prod_{j=1}^{m} \gamma_j^{I_j(z_i)} * [\beta|\delta] * [\gamma]. \end{aligned} \tag{2.6}$$

So to apply the Gibbs sampler, the desired full conditional distributions are as follows:
For $1 \le i \le l(Z_1, ..., Z_l \; or \; X_{n-l+1}, ..., X_n)$,

$$[Z_i|Y, X_k, 1 \le k \le n-l, Z_j, 1 \le j \le l, j \ne i, \beta, \gamma]$$
$$= \frac{exp\left\{Y_{n-l+i}[\beta_0 + \sum_{j=1}^{p-1} Z_{ij}\beta_j] - log[1+exp(\beta_0+\sum_{j=1}^{p-1} Z_{ij}\beta_j)]\right\}\prod_{j=1}^{m}\gamma_j^{I_j(Z_i)}}{\sum exp\left\{Y_{n-l+i}[\beta_0 + \sum_{j=1}^{p-1} Z_{ij}\beta_j] - log[1+exp(\beta_0+\sum_{j=1}^{p-1} Z_{ij}\beta_j)]\right\}\prod_{j=1}^{m}\gamma_j^{I_j(Z_i)}}, \tag{2.7}$$

where the summation is over all cases of X's (In this case, there are m summations).

Let $(\Sigma^{-1})_{ij} = \tau_{ij}$, $\mu_i' = \mu_i - \frac{1}{\tau_{ij}}\sum_{j\ne i}(\beta_j - \mu_j)\tau_{ij}$ for $i = 0, 1, \ldots, p-1$ and $j = 0, 1, \ldots, p-1$.

$$\begin{aligned}&[\beta_0|Y, X_k, 1 \le k \le n-l, Z_i, 1 \le i \le l, \beta_j, j \ne 0, \gamma] \\ &\propto exp\left\{\sum_{i=1}^{n}[y_i\beta_0 - log(1+exp(\beta_0+\sum_{j=1}^{p-1} x_{ij}\beta_j))] - \frac{\tau_{00}}{2}(\beta_0-\mu_0')^2\right\}\end{aligned} \tag{2.8}$$

for $1 \le k \le p-1$,

$$\begin{aligned}&[\beta_k|Y, X_k, 1 \le k \le n-l, Z_i, 1 \le i \le l, \beta_j, j \ne k, \gamma] \\ &\propto exp\left\{\sum_{i=1}^{n}[y_i x_{ik}\beta_k - log(1+exp(\beta_0+\sum_{j=1}^{p-1} x_{ij}\beta_j))] - \frac{\tau_{kk}}{2}(\beta_k-\mu_k')^2\right\}\end{aligned} \tag{2.9}$$

and since $[\gamma|Y, X_k, 1 \leq k \leq n-l, Z_i, 1 \leq i \leq l, \beta] \propto [\gamma] * \prod_{i=1}^{l}\prod_{j=1}^{m}\gamma_j^{I_j(z_i)}$, for $1 \leq k \leq m$,

$$[\gamma_k|Y, X, \beta, \gamma_j, j \neq k] \propto \gamma_k^{\alpha_k - 1}\left(1 - \sum_{i=1}^{m}\gamma_i\right)^{\alpha_{m+1}-1}\prod_{i=1}^{l}\prod_{j=1}^{m}\gamma_j^{I_j(z_i)}. \quad (2.10)$$

2.2. Probit model

Under the usual probit model, for $i = 1, \ldots, n$, let

$$\Phi^{-1}(p_i) = \beta_0 + X_{i1}\beta_1 + \ldots + X_{ip-1}\beta_{p-1}.$$

Using the idea of Albert and Chib (1993), define the latent variables $W = (W_1, \ldots, W_n)$ such that $W_i \sim N(X_i{}^t\beta, 1)$, and such that $W_i \leq 0 \Leftrightarrow Y_i = 0$, and $W_i > 0 \Leftrightarrow Y_i = 1$.

Now the joint probability density for Y, W, β, δ and γ is

$$\begin{aligned}[Y, W, \beta, \delta, \gamma|X] &= [W|X, \beta][\beta|\delta][\gamma][Y|W, \beta] = [W|X, \beta][\beta|\delta][\gamma][Y|W] \\ &= [W|X, \beta][\beta|\delta][\gamma] \\ &= exp\{-\frac{1}{2}\sum_{i=1}^{n}(w_i - \tilde{x}_i{}^t\beta)^2\} * \prod_{i=1}^{l}\prod_{j=1}^{m}\gamma_j^{I_j(z_i)}[\beta|\delta][\gamma]. \quad (2.11)\end{aligned}$$

For $1 \leq i \leq l(Z_1, ..., Z_l \; or \; X_{n-l+1}, ..., X_n)$,

$$\begin{aligned}&[Z_i|Y, W, X_k, 1 \leq k \leq n-l, Z_j, 1 \leq j \leq l, j \neq i, \gamma, \beta] \\ &= \frac{exp\left\{-\frac{1}{2}(w_i - (\beta_0 + \sum_{j=1}^{p-1}x_{ij}\beta_j))^2\right\}\prod_{j=1}^{m}\gamma_j^{I_j(Z_i)}}{\sum exp\left\{-\frac{1}{2}(w_i - (\beta_0 + \sum_{j=1}^{p-1}x_{ij}\beta_j))^2\right\}\prod_{j=1}^{m}\gamma_j^{I_j(Z_i)}} \quad (2.12)\end{aligned}$$

where the summation is over all cases of X's.

$$\begin{aligned}&[\beta_0|Y, W, X, \beta_j, j \neq 0, \gamma] \\ &= N\left(\frac{\tau_{00}\mu_0' + \sum_{i=1}^{n}(w_i - \sum_{j=1}^{p-1}x_{ij}\beta_j)}{n + \tau_{00}}, (n + \tau_{00})^{-1}\right) \quad (2.13)\end{aligned}$$

and for $1 \leq k \leq p-1$,

$$\begin{aligned}&[\beta_k|Y, W, X, \beta_j, 1 \leq j \leq p-1, j \neq k, \gamma] \quad (2.14) \\ &= N\left(\frac{\tau_{kk}\mu_k' + \sum_{i=1}^{n}x_{ik}[w_i - \sum_{j\neq k}x_{ij}\beta_j - \beta_0]}{\sum_{i=1}^{n}x_{ik}^2 + \tau_{kk}}, (\sum_{i=1}^{n}x_{ik}^2 + \tau_{kk})^{-1}\right)\end{aligned}$$

where $N(a,b)$ denotes the normal density with mean a and variance b.
Finally, for $1 \leq k \leq m$, $[\gamma_k|Y,X,\beta,\gamma_j,j \neq k]$ is in the form of (2.10).
The implemetation of Gibbs sampler is briefly described in the following.

Step 1. Starting with initial guesses at $\beta_0^{(0)},\beta_1^{(0)},\dots,\beta_{p-1}^{(0)}$, $Z_1^{(0)}$, ..., $Z_l^{(0)}$, $\gamma_1^{(0)}$, ..., $\gamma_m^{(0)}$, simulate the $W^{(1)}$ from the truncated normal distribution.

Step 2. The usual Gibbs iteration is as folows:

For $i = 1,\dots,l$,

$$z_i^{(1)} \sim [z_i|Y,W^{(1)},X_k,1 \leq k \leq n-l,\gamma^{(0)},\beta^{(0)},Z_j = z_j^{(0)}\ for\ j > i\ and\ z_j^{(1)}\ for\ j < i]$$

and for $k = 0,\dots,p-1$,

$$\beta_k^{(1)} \sim [\beta_k|Y,W^{(1)},X_k,1 \leq k \leq n-l,Z = z^{(1)},\beta_j^{(0)}\ for\ j > k\ and\ \beta_j^{(1)}\ for\ j < k,\gamma^{(0)}]$$

and for $k = 1,\dots,m$,

$$\gamma_k^{(1)} \sim [\gamma_k|Y,W^{(1)},X_k,1 \leq k \leq n-l,Z = z^{(1)},\beta^{(1)},\gamma_j^{(0)}\ for\ j > k\ and\ \gamma_j^{(1)}\ for\ j < k].$$

The above two steps form an iteration which updates $\beta^{(0)},\gamma^{(0)},Z^{(0)},W^{(0)}$ to $\beta^{(1)},\gamma^{(1)},Z^{(1)}$, $W^{(1)}$. Thus t such iterations produce a "one-string run". Also, n parallel strings are run with different starting positions to make sure that the samples converge to the whole posterior distribution, instead of a local maximum of the posterior distribution. For the logit model, $W^{(1)}$ from the truncated normal in the step 1 is not needed.

2.3. Estimating the marginal posterior distributions

Let $\beta = (\beta_0,\dots,\beta_{p-1})$ be a parameter vector and let Z latent data possibly. Let $\{(\beta_0^{(g)},\dots,\beta_{p-1}^{(g)}),z^{(g)}\}_{g=1}^G$ be sample from posterior, that is, this Gibbs output can be drawn from Gibbs sampler with full conditional densities in (2.7) - (2.10) or (2.12) - (2.14).

For the probit model, since the full conditional densities in (2.13) and (2.14) are in closed form, by Rao-Blackwellized estimation , $\pi(\beta_i|Y)$ can be estimated by

$$\hat{\pi}(\beta_i|y) = \frac{1}{G}\sum_{g=1}^{G}\pi(\beta_i|\beta_j^{(g)},j \neq i,z^{(g)}). \tag{2.15}$$

But for the logit model, we can not use the Rao-Blackwellized estimation in (2.15) because the full conditional densities in (2.8) and (2.9) are not in closed form. Then, Gelfand and Smith(1990) suggested any smoothed kernel density estimator of $\pi(\beta_i|Y)$ based on the Gibbs output $\{\beta_i^{(g)}\}_{g=1}^G$ for any estimated marginal density of $\pi(\beta_i|Y)$. In this time, without using kernel density estimation, we estimate the posterior marginal density $\pi(\beta_i|y)$ using the Monte Carlo method. The idea is very simple using the fact that $\pi(\beta|y)m(y) = l(\beta|y)\pi(\beta)$ where $l(\beta|y)$, $m(y)$ and $\pi(\beta)$ are the sampling density, the marginal density of y and the (possibly improper) prior density of β, respectively. Let $g(\beta) = l(\beta|y)\pi(\beta)$ and let $q(\beta)$ be arbitrary density function.

We want to estimate the posterior marginal density, say $\pi(\beta_i|Y)$. First, we evaluate the value of $\pi(\beta_i|Y)$ at each point β_i^*. For fixed β_i^*,

$$\begin{aligned}
\pi(\beta_i^*|Y) &= \int \pi(\beta_i^*, \beta_{[-i]}|Y)d\beta_{[-i]} \\
&= \int \pi(\beta_i^*, \beta_{[-i]}|Y) \int q(\beta_i|\beta_{[-i]})d\beta_i d\beta_{[-i]} \\
&= \int\int q(\beta_i|\beta_{[-i]})\frac{\pi(\beta_i^*, \beta_{[-i]}|Y)}{\pi(\beta_i, \beta_{[-i]}|Y)}\pi(\beta_i, \beta_{[-i]}|Y)d\beta_i d\beta_{[-i]} \\
&= \int\int q(\beta_i|\beta_{[-i]})\frac{g(\beta_i^*, \beta_{[-i]})}{g(\beta_i, \beta_{[-i]})}\pi(\beta_i, \beta_{[-i]}|Y)d\beta_i d\beta_{[-i]} \\
&\sim \frac{1}{G}\sum_{g=1}^{G} q(\beta_i^{(g)}|\beta_{[-i]}^{(g)})\frac{g(\beta_i^*, \beta_{[-i]}^{(g)})}{g(\beta_i^{(g)}, \beta_{[-i]}^{(g)})} \qquad (2.16)
\end{aligned}$$

where $\beta_{[-i]} = (\beta_0, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_{p-1})$ and $q(\beta_i|\beta_{[-i]})$ is conditional density of β_i obtained from $q(\beta_0, \ldots, \beta_i, \ldots, \beta_{p-1})$ and $\{\beta^{(g)}\}_{g=1}^G$ is Gibbs output.

The grid points β_i^* need not to be uniformly spaced. But the number of points in the grid need be as large as possible on the support line. Choosing a good function q can be quite difficult. In some cases a reasonable choice of q is to use a normal density whose mean and variance are based on the sample mean and sample covariance of Gibbs output $\{\beta^{(1)}, ..., \beta^{(g)}\}$. Verdinelli and Wasserman (1995) used the similar idea when computing the Bayes factor.

3. BAYESIAN NODEL CHOICE

In this section, we test the proposed model v.s. the model deleting the incomplete data(see Vach and Schumacher, 1993) and also the logit and probit models are compared using Bayes factor. In general, suppose that we are interested in comparing two models M_0 and M_1. The formal Bayesian model choice procedure goes as follows. Let w_i be the prior probability of M_i, $i = 0, 1$ and let $f(y|M_i)$ be the predictive distribution for model M_i, i.e.

$$[m]_i = f(y|M_i) = \int f(y|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i \ .$$

If y is the observed data, then we choose the model yielding the larger $w_i f(y|M_i)$. Often we set $w_i = \frac{1}{2}$ and compute the Bayes factor (or M_0 with respect to M_1)

$$BF = \frac{f(y|M_0)}{f(y|M_1)} = \frac{[m]_0}{[m]_1}. \tag{3.1}$$

Jeffreys(1961) and Kass and Raftery(1995) suggest interpretive ranges for the Bayes factor and in general , M_0 is supported if $BF > 1$.

More generally, we want to estimate $[m] = \int f(y|\beta)\pi(\beta)d\beta$ using the importance sampling method. Let us consider $\pi(\beta|y)$ as the importance sampling function. Then the Markov Chain Monte Carlo methods, particularly Metropolis algorithm and Gibbs sampler, are used to get the sample from the posterior density $\pi(\beta|y)$. Let $\{\beta^{(g)}\}_{g=1}^{G}$ be Gibbs outputs as above. Then by Monte Carlo method, the approximating marginal density of Y is $\hat{[m]} = \frac{\sum_{g=1}^{G} w_g f(y|\beta^{(g)})}{\sum_{g=1}^{G} w_g}$ where $w_g = \frac{\pi(\beta^{(g)})}{\pi(\beta^{(g)}|y)}$. Since $\pi(\beta|y) = \frac{f(y|\beta)\pi(\beta)}{[m]}$, the approximation can be expressed as

$$\hat{[m]} = \left[\frac{1}{G}\sum_{g=1}^{G}\frac{1}{f(y|\beta^{(g)})}\right]^{-1}. \tag{3.2}$$

Also, this final form is mentioned in Kass and Raftery(1995).

For example, suppose that we want to test the logit model (M_0) v.s. the probit model (M_1), then the approximating Bayes factor for favoring M_0 is given by

$$BF = \frac{[\frac{1}{G}\sum_{g=1}^{G}\{\prod_{i=1}^{N} L(\tilde{x}^t\beta^{(g)})^{y_i}(1 - L(\tilde{x}^t\beta^{(g)}))^{1-y_i}\}^{-1}]^{-1}}{[\frac{1}{G}\sum_{g=1}^{G}\{\prod_{i=1}^{N} \Pi(\tilde{x}^t\beta^{(g)})^{y_i}(1 - \Pi(\tilde{x}^t\beta^{(g)}))^{1-y_i}\}^{-1}]^{-1}} \tag{3.3}$$

where $L(.)$ and $\Pi(.)$ denote the logistic cdf and standard Gaussian cdf, respectively. In particular, for the case containing the missing data, they are replaced by the Gibbs outputs corresponding to them into the approximating BF in (3.3).

4. ILLUSTRATIVE EXAMPLE

In this section, we shall consider Bayesian analysis, using the Gibbs sampler, of logistic model. Ibrahim(1990) considered the study of 82 patients who experienced translaryngeal intubation(TLI) for more than four days and were prospectively evaluated for laryngeal complications in Table 4.1. At the begining of study, data was collected on the patients regarding 13 baseline explanatory variables (covariates) during the period of TLI. The reponse variable (y) is dichotomized to 0 or 1, with 0 and 1 representing no damage and damage of the larynx at baseline, respectively. Ibrahim(1990) used only three covariates in his model, which correspond to the covariates with the largest fraction of missing data. These three covariates consist of serum albumin (x_1), serum creatinine (x_2) and the third variable (x_3) which is the ratio of laryngeal size to tracheal tube size. These are all dichotomous. Of 82 patients, there are 13 patients who have no answer as missing data. For this data, Ibrahim(1990) introduced the method to find MLE in logistic model with missing data using EM algorithm. Also we consider the regression model $\beta_0 + \sum_{i=1}^{3} \beta_i x_i$ used by Ibrahim(1990).

Table 4.1 TLI data

x_1	0	0	0	0	1	1	0	0	0	0	1	1	1	.	0	1	.	.	0	0	1	1
x_2	0	0	1	1	0	0	0	0	1	1	0	0	1	0	1	1	.	0	.	0	0	1
x_3	0	1	0	1	1	0	0	1	0	1	0	1	0	0	.	.	.	.	0	.	.	.
y	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1
F	10	10	6	2	7	1	9	10	2	7	3	1	1	1	1	1	1	1	1	2	1	4

where $\cdot$ and "F" denote a missing data and frequency, respectively.

In this case, $N = 82, l = 13, p = 4$ and $m = 8$. Assume that the prior of β is noninformative, i.e. $\mu = 1$, $\Sigma^{-1} = 0$ and set $\alpha_i = 1$ for $i = 1, \ldots, 8$ in (2.7). For logit model, the Metropolis algorithm is applied to (2.8), (2.9) and (2.10). But, for probit model, the Metropolis algorithm is only applied to (2.10) because (2.13) and (2.14) have the explicit forms of their density functions. For these cases using Metropolis algorithm, normal density function is used as the derived function. Each convergence of Gibbs sampler and Metropolis algorithm is checked using Gelman and Rubin's(1992) method.

In Table 4.2, CD and CR denote the estimates of the posterior means of β_i's when deleting missing data and replacing missing data by random variates, respectively. Figures 4.1 and 4.2 indicate the estimated densities of β in the logit and probit models, respectively. In particular, the graphs in figure 4.1 are sketched by the estimates in (2.16) after getting the normalizing constants. Figure 4.2 is graphed by the Rao-Blackwellized estimates in (2.15) since the full conditional densities of β_i's are fully known. In Figures 4.1 and 4.2, the graphs of $\beta_0, \beta_1, \beta_2$ and β_3 are denoted by solid, dotted, dashed and long-dashed lines, respectively.

Next, we consider the logit model and then test the model M_0 v.s. the model M_1 as follows;

M_0: the model $f(y|\beta)$ replacing the incomplete data with the new variates using data augmentation with normal prior.

v.s.

M_1:the model $f(y|\beta)$ deleting the incomplete data with normal prior.

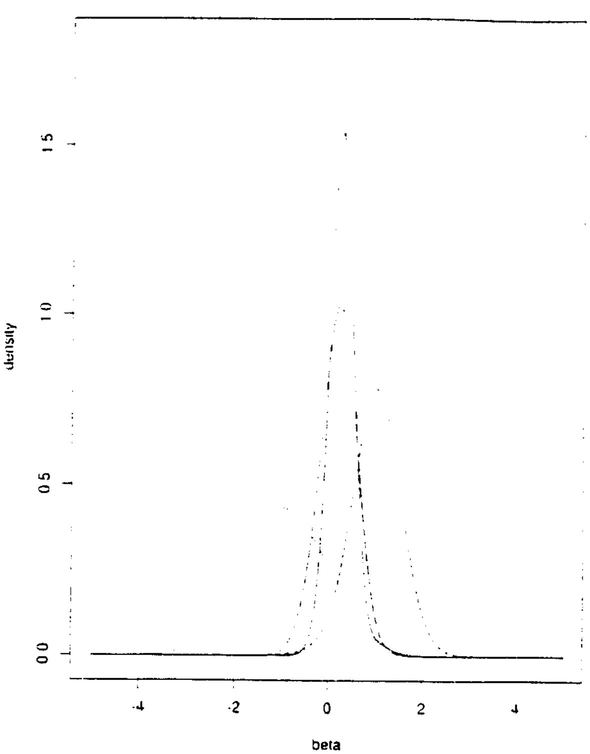

Figure 4.1 logit model

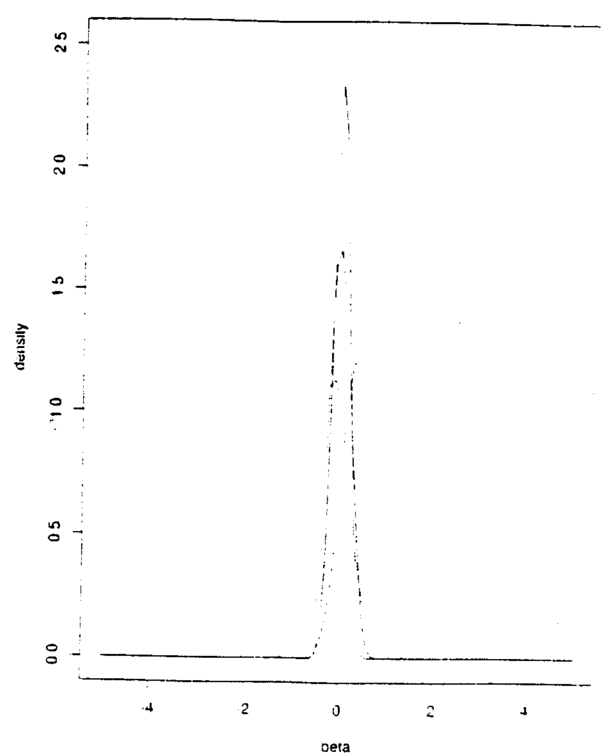

Figure 4.2 probit model

Then the Bayes factor for favoring M_0 is 2.3 x 10^2 and so for the given data, the method M_0 is appropriate. Similarly for probit model, the approximate

Bayes factor for favoring M_0 is 3.1 x 10^3. Finally, we compare the logit model(M_0) with the probit model(M_1) and then the Bayes factor given by (3.3) is 4.1 x 10^3 and so, the logit model is more appropriate than the probit model for the given data.

Table 4.2 posterior Mean

	Logit		Probit	
	CD	CR	CD	CR
β_0	-0.1588	0.5308	-0.0581	-0.1000
(S.E.)	(0.2933)	(0.0661)	(0.2241)	(0.2511)
β_1	-1.3677	-1.0382	-0.3789	0.0739
(S.E.)	(0.2271)	(0.2948)	(0.3622)	(0.4219)
β_2	1.1691	0.8415	0.4135	0.3879
(S.E.)	(0.2172)	(0.2366)	(0.3739)	(0.3643)
β_3	-0.1403	0.4964	-0.0485	0.0198
(S.E.)	(0.1546)	(0.1172)	(0.3162)	(0.2756)

REFERENCES

(1) Albert, J. and Chib, S. (1993) Bayesian Analysis of binary and Polychotomous response data, *Journal of American Statistical Association*, **88**, 669-679.

(2) Dellaportas, P. and Smith, A.M.F. (1993) Bayesian Inference of Generalized Linear Model via Gibbs Sampling, *Applied Statistics*, **42**, 443-459.

(3) Dempster, A.P., Laird, N. and Rubin, D.B. (1977) Maximum Likelihood From Incomplete Data via the EM algorihm, *Journal of Royal Statistical Society, Series B.*, **39**, 1-38.

(4) Fuchs, C. (1982) Maximum Likelihood Estimation and Model Selection in Contigency Tables with missing data, *Journal of American Statistical Association*, **77**, 270-278.

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

(6) Gelman, A. and Rubin, D.B.(1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, No. 4, 457-472.

(7) Ibrahim, J.G. (1990) Incomplete data in Generalized Linear models, *Journal of American Statistical Association*, **85**, 765-769.

(8) Kass E. K. and Raftery A. E. (1995) Bayes Factors *Journal American Statistical Association*, **90**, 773-795.

(9) McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear models*, 2nd ed. Londen : Champan and Hall.

(10) Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller,A.H., and Teller,E.(1953) Equation of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, **21**, 1087-1092.

(11) Stephens, D.A. and Dellaportas, P. (1992) *Bayesian Analysis of Generalized Linear models with Covariate Measurement error*, Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.) Oxford, University press, 813-820.

(12) Tanner, M. and Wong, W.H. (1987) The calculation of posterior distributions by data augmentation(with discussion), *Journal of American Statistical Association*, **82**, 528-550.

(13) Vach, W. and Schumacher, M. (1993) Logistic regression with incmpletely observed categorical covariates: A comparison of three approaches, *Biometrika*, **80**, 353-362.

(14) Verdinelli I. and Wasserman L. (1995) Computing Bayes factors using a generalization of the Savege-Dickey density ratio , *Journal of American Statistical Association*, **90**, 614-618.