Hacettepe Journal of Mathematics and Statistics
Volume 37(2) (2008), 167–176

A BAYESIAN APPROACH TO PARAMETER ESTIMATION IN BINARY LOGIT AND PROBIT MODELS

Derya Tektaş* and Süleyman Günay*

Received 07:03:2006 : Accepted 16:10:2008

Abstract

In the areas of statistics and econometrics the analysis of binary and polychotomous response data is widely used. In classical statistics, the maximum likelihood method is used to model this data and inferences about the model are based on the associated asymptotic theory. However, the inferences based on the classical approach are not accurate if the sample size is small. J.H. Albert and S. Chib (*Bayesian Analysis of Binary and Polychotomous Response Data*, J. American Statistical Association **422**, 669–679, 1993) proposed a Bayesian method to model categorical response data. In this method Gibbs sampling and the data augmentation algorithm are used together to model the data. In this article, Albert and Chib's approach is used to estimate the parameters in the logit and probit models. Furthermore, the maximum likelihood and ordinary least-squares methods are discussed briefly, and a simple example is presented to compare these three methods.

Keywords: Binary probit model, Binary logit model, Bayesian analysis, Albert and Chib approach, Latent data, Gibbs sampling, Data augmentation.

2000 AMS Classification: 62J12, 62F5, 65C05.

1. Introduction

If the dependent variable in a data set is categorical, Generalized Linear Models (GLMs) are used instead of linear regression models to estimate the model parameters and to model the data. GLMs can be expresed in the form $\mathrm{E}(y) = g(x'\beta)$, where y is a response variable which is categorical, x is the vector of explanatory variables, β is the vector of model parameters and g is the link function. GLMs depend on a probability model that describes an event's probability as the distribution function of the independent variables. The model is binary if the dependent variable takes two

*Department of Statistics, Faculty of Science, Hacettepe University, 06800 Beytepe, Ankara, Turkey. E-mail : (D. Tektaş) dtektas@hacettepe.edu.tr (S. Günay) sgunay@hacettepe.edu.tr

values ($y = 0, 1$). However, the model is polychotomous or multinomial if the dependent variable takes more than two values.

In the Bayesian analysis of linear regression models conjugate prior distributions exist for the model parameters. Thus, it is easy to make inferences for these models. However, making inferences in Bayesian GLMs is complicated because there are no conjugate prior distributions for the model parameters. So, it is hard to make a simulation in Bayesian GLMs. Albert and Chib [1] proposed an approach for binary probit regression models to resolve this problem. In this approach the conditional distributions of the model parameters in GLMs become the same as those in linear regression models by adding a latent variable to the model [4].

In this article a simulation based approach proposed by Albert and Chib [1] is introduced for computing the exact posterior distribution of β. In this approach, the data set in the logit and probit models can be augmented by adding a set of latent variables (z) into the model. Latent variables have a continuous distribution such that $y_i = 1$ when $z_i > 0$ and $y_i = 0$ otherwise. Thus, the conditional distribution of β given z is a normal distribution whose mean can be easily computed, and the conditional distribution of z given β is a truncated normal distribution which is easy to compute. Since the conditional distributions can be calculated using this approach, Gibbs sampling can easily be used to calculate the exact posterior distribution of β.

This article is organized as follows. Section 2 outlines the Gibbs sampling and data augmentation algorithm used in simulating the posterior distributions. Section 3 discuses binary regression models based on latent data. In Section 4, an example using the data set from the Econometric Toolbox by LeSage [5] is given. In this example, to estimate the parameters in the logit and probit models, the maximum likelihood (ML), ordinary least-squares (OLSs) and Bayesian methods are used and the results are compared. In addition, the Bayesian, logit and probit models are compared according to their convergence behaviours. Finally, Section 5 presents some concluding remarks.

2. Gibbs sampling and the data augmentation algorithm

Gibbs sampling and the data augmentation algorithm are both used to compute the posterior distribution of β. Albert and Chib [1] used these two methods together to analyze logit and probit models.

Gibbs sampling allows one to sample from a multivariate distribution using full conditional distributions. A full conditional distribution is the conditional distribution of a parameter given all of the other parameters in the model. The data set which is obtained by applying Gibbs sampling to the model converges to the joint posterior distribution of the parameters [2, 5].

For the parameters β_i ($i = 1, \dots, p$) the initial values are taken as $\beta_1^{(0)}, \dots, \beta_p^{(0)}$. After the initial values of the parameters are determined, Gibbs sampling is implemented by sampling from the full conditional distributions successively.

$$(1) \qquad \begin{array}{l} \beta_1^{(1)} \text{ from } \pi\big(\beta_1/\{\beta_j^{(0)},\ j \neq 1\}\big) \\ \beta_2^{(1)} \text{ from } \pi\big(\beta_2/\beta_1^{(1)}, \{\beta_j^{(0)},\ j > 2\}\big) \\ \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \beta_p^{(1)} \text{ from } \pi\big(\beta_p/\{\beta_j^{(1)},\ j < p\}\big) \end{array}$$

The cycle (1) is iterated t times and so the sample $\beta(t) = (\beta_1^{(t)}, \dots, \beta_p^{(t)})$ is generated. As t approaches infinity, the joint distribution of $\beta(t)$ approaches the joint distiribution of β. In applications, it is necessary to specify how many iterations will be sufficient. If this number is t^*, $\beta(t^*)$ can be described as a value simulated from the posterior distribution of β.

After t^* is determined, the draws for $t < t^*$ are discarded from the sample. Afterwards, the sample $\{(\beta_{1j}^{(t^*)}, \beta_{2j}^{(t^*)}, \ldots, \beta_{pj}^{(t^*)}),\ j = 1, \ldots, m\}$ is produced by replicating this process m times. This sample can be used to to compute estimations for the posterior moments and density. In this approach as proposed by Albert and Chib [1], one replication is used and the cycle (1) is run a sufficient number of times to produce convergence.

The main objective of Gibbs sampling is to estimate the marginal posterior distributions of subsets of β, and to derive posterior moments by simulating a sufficiently large number of values from the joint posterior distribution of β. It is possible to obtain all the characteristics of marginal posterior distributions by using the Gibbs sequence. For example; we can estimate the function $g(\beta_k)$ with the Gibbs sequence of β_k. The simulated values of this function can be derived by using the simulated values of β_k. Thus, a density estimate of this function can be obtained by using kernel density estimates of simulated values of $g(\beta_k)\{g(\beta_k^{(i)}),\ i = 1, \ldots, m\}$. According to Gelfand and Smith [2], Equation (2) can be used to find accurate estimate of this marginal posterior density [1].

$$(2) \qquad \widehat{\pi}\big(g(\beta_k)\big) \approx \frac{1}{m}\sum_{i=1}^{m} \pi\big(g(\beta_k)/\{\beta_r^{(i)},\ r \neq k\}\big).$$

A difficulty with Gibbs sampling is that the sequences produced by this method are not independently identically distributed. Besides, convergence of the sequence to the posterior distribution is an important point in Gibbs sampling. Theoretically, as the number of draws (n) in a Gibbs sequence approaches to infinity, the sequence converges to the posterior distribution. However, a sufficient number of draws for convergence of the sequence must be determined in practice.

There are many diagnostics for Gibbs sampling to determine whether the sequence converges to the posterior distribution or not. Some of these convergence diagnostics are autocorrelation estimates, Raftery-Lewis diagnostics, Geweke numerical standard errors (NSE) and relative numerical efficiency (RNE) estimates, Geweke Chi-squared test on the means from the first 20% of the sample versus the last 50%.

Autocorrelation estimates show how much independence exists among the draws in each β parameter sequence. If the autocorrelation among draws in a parameter sequence is high, both mixing and convergence of this sequence is slow. So, it is recommended to increase the thinning interval to reduce the autocorrelation [5].

Besides the above, Raftery and Lewis [6] proposed some convergence diagnostics. The method proposed by Raftery and Lewis [6] is easy to implement and depends on quantiles of functionals of the posterior distribution. In this method the total number of iterations required and the thinning interval, which is an indicator of autocorrelation among the draws, are determined. In addition to this, the number of initial iterations which should be discarded from the parameter sequence is calculated [5, 6].

Moreover, the diagnostics proposed by Geweke [3] are used to determine the amount of autocorrelation among the draws in a parameter sequence, and whether the parameter sequence has reached an equilibrium state. These diagnostics are based on estimates of the numerical standard error (NSE) and relative numerical efficiency (RNE) [3, 5].

The second method besides Gibbs sampling which is used to compute the posterior distributions is the data augmentation algorithm. In this algorithm, the observed data y is augmented by latent data z. Therefore, it is easy to analyze the observed data y. To implement this algorithm, samples are drawn from the conditional distributions $p(\beta/y, z)$ and $p(z/\beta, y)$. The posterior distribution in Equation (3) is derived by implementing this algorithm [8].

$$p(\beta/y) = \int_z p(\beta/z, y)p(z/y)dz. \tag{3}$$

To compute the posterior density in Equation (3), the predictive density $p(z/y)$ of z must be computed using Equation (4).

$$p(z/y) = \int_B p(z/\beta, y)p(\beta/y)d\beta. \tag{4}$$

The data augmentation algorithm to compute the posterior density of β can be implemented by applying the following steps. In the ith step of the algorithm, the current approximation to $p(\beta/y)$ is taken to be $g_i(\beta)$,

(a) Draw a sample $z^{(1)}, \ldots, z^{(m)}$ from the current approximation to the predictive density $p(z/y)$.

(b) Update the current approximation to $p(\beta/y)$ to give $g_{i+1}(\beta)$ as a combination of conditional densities of β given the augmented data which is generated in step (a). Specifically, $g_{i+1}(\beta)$ is computed as given below.

$$g_{i+1}(\beta) = m^{-1} \sum_{j=1}^{m} p(\beta/z^j, y).$$

In steps (a_1) and (a_2) below, it is shown how the latent data z is generated from $p(z/y)$ in step (a). Equation (4) allows us to generate z from $p(z/y)$ in these steps when the current approximation to $p(\beta/y)$ is $g_i(\beta)$.

(a_1) Generate β from $g_i(\beta)$

(a_2) Generate z from $p(z/y)$ (β is the value generated in step (a_1)).

If the number of draws (m) is sufficiently large, a good approximation to $p(\beta/y)$ can be achieved by implementing the steps (a_1), (a_2), (a) and (b) [7, 8].

3. Bayesian analysis of the binary Logit and Probit models

Some representations for binary regression models are given below.

$$\begin{aligned} &y_i \sim \text{Bernoulli}(g(\eta_i)) \\ &\eta_i \sim x_i'\beta \\ &\beta \sim \pi(\beta). \end{aligned} \tag{5}$$

In the representation (5), y_i is a binary response variable that takes only two values 0 and 1 for n observations; $x_i' = (x_{i1}, \ldots, x_{ip})$ are p covariate measurements; β is a $(p \times 1)$ vector of regression coefficients and $\pi(\cdot)$ is the prior distribution of β. Here $g(\cdot)$ is a link function and η_i a linear predictor. If the link function is the standard Gaussian cumulative distribution function (cdf), a probit model is obtained. If the link function is the logistic cdf, the logit model is obtained. [1, 4].

The Ordinary Least Squares (OLSs) method can be used to analyze the binary regression model, but using this method with a binary response variable causes some problems. One of the problems is that the errors are heteroscedastic and these heteroscedastic errors are a function of the parameter vector β. Another problem is that the predicted values can take values outside the interval $(0, 1)$ when the regression model is analyzed with OLSs method [5].

The Maximum Likelihood (ML) method can also be used to analyze binary response regression models. If $\pi(\beta)$ is a proper or improper prior density for β, the posterior density of β can be calculated using Equation (6).

$$\pi(\beta/y) = \frac{\pi(\beta)\prod_{i=1}^{N} g(x_i'\beta)^{y_i}(1-g(x_i'\beta))^{1-y_i}}{\int_{\mathrm{B}} \pi(\beta)\prod_{i=1}^{N} g(x_i'\beta)^{y_i}(1-g(x_i'\beta))^{1-y_i}\,d\beta}. \tag{6}$$

A Multivariate normal distribution with p variables can be denoted by $N_p(\mu, \Sigma)$, where μ is the mean vector and Σ is the variance-covariance matrix. According to the usual asymptotic approximation, the distribution of β is $N_p(\widehat{\beta}, I(\widehat{\beta})^{-1})$. In this notation, $\widehat{\beta}$ denotes the posterior mode and $I(\widehat{\beta})$ the negative of the second derivative matrix evaluated at $\widehat{\beta}$. If the prior distribution of β is uniform, $\hat{\beta}$ and $I(\cdot)$ can be defined as the maximum likelihood estimate (MLE) of β and the observed information matrix respectively. However, MLEs are biased for small samples. Hence, it is not accurate to use a normal approximation for small samples [1].

As mentioned in Section 1, Albert and Chib [1] proposed a Bayesian method to analyze regression models where the response variable is categorical. This approach depends on the idea that the posterior distribution of β given the latent variable z in categorical regression models, and the posterior distribution of β in normal linear regression models ($z = x'\beta + \varepsilon$) are the same.

Let the model represented in Equation (5) be analyzed by this Bayesian method for the probit link. After the latent variables are added to the model, it can be represented as given below [4].

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise.} \end{cases} \\ z_i &= x_i'\beta + \varepsilon_i,\ \varepsilon_i = N(0,1),\ \beta \sim \pi(\beta) \end{aligned} \tag{7}$$

It is easy to find the full conditional distributions of β and z by using representation (7). Since these full conditional distributions can be calculated, Gibbs sampling can easily be performed to analyze the binary probit model.

If the prior distribution of β is normal, that is $\pi(\beta) = N(b, v)$, the full conditional distribution of β is also normal.

$$\beta/z \sim N(B, V),\ B = V(v^{-1}b + x'z),\ V = (v^{-1} + x'x)^{-1} \tag{8}$$

If β has a diffuse prior distribution, then the full conditional distribution of β is also normal.

$$\beta/y, z \sim N(\widehat{\beta}_Z, (x'x)^{-1}), \tag{9}$$

where $\hat{\beta}_z = (x'x)^{-1}(x'z)$.

The full conditional distribution of each latent variable z_i is a truncated normal distribution which is easy to simulate.

$$z_i/\beta, x_i, y_i \sim \begin{cases} N(x_i'\beta, 1)I(z_i > 0) & \text{if } y_i = 1, \\ N(x_i'\beta, 1)I(z_i \le 0) & \text{otherwise.} \end{cases} \tag{10}$$

In practice β usually has a flat noninformative prior distribution. When a previous value of β is given, a cycle of Gibbs sampling simulates z and β from the distributions (10) and (9), respectively. The initial value $\beta^{(0)}$ of β can be determined as ML or OLS estimates of β. Sampling from the distributions (9) and (10) is straightforward [1, 5].

The method proposed by Albert and Chib [1] to analyze binary probit models can be developed by generalizing the probit link function as a family of t distributions. Thus, the degree of freedom of the t distribution which is best suited to the data, can be determined. The most popular link functions for binary data are logit and probit. If the degree of freedom is taken as seven, then the link function is logit. Moreover if the degree of freedom is taken to be a very large number (in practice it is taken to be 100), then the link function is probit.

If ε_i has a standard logistic distribution in the representation (7), the binary logit model is obtained. When the variables $\lambda_i,\ i = 1,\dots,n$ are introduced into the model, then the binary logit model can be represented as in (11).

$$(11)\qquad \begin{aligned} y_i &= \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise.} \end{cases} \\ z_i &= x_i'\beta + \varepsilon_i,\ \varepsilon_i \sim N(0,\lambda_i),\ \lambda_i = (2\psi_i)^2,\ \psi_i \sim K,\ \beta \sim \pi(\beta), \end{aligned}$$

where the random variables $\psi_i,\ i = 1,\dots,n$ are independent and have the Kolmogorov-Smirnov (KS) distribution [4].

If β has a normal prior distribution, $\pi(\beta) = N(b,v)$ in the binary logit model, then the full conditional distribution of β given z, y and λ is normal as represented in (12).

$$(12)\qquad \begin{aligned} &\beta/y,z,\lambda \sim N(B,V), \\ &B = V(v^{-1}b + x'Wz),\ V = (v^{-1} + x'Wx)^{-1},\ W = diag(\lambda_1^{-1},\dots,\lambda_n^{-1}). \end{aligned}$$

If the prior distribution of β is uniform, then the full conditional distribution of β is normal.

$$(13)\qquad \beta/y,Z,\lambda \sim N(\hat{\beta}_{z,\lambda},(x'Wx))$$

where $\widehat{\beta}_{z,\lambda} = (x'Wx)^{-1}x'Wz$.

The full conditional distribution of each latent variable z_i is truncated normal with variance λ_i.

$$(14)\qquad z_i/\beta,x_i,y_i,\lambda_i \sim \begin{cases} N(x_i'\beta,\lambda_i)I(z_i > 0) & \text{if } y_i = 1, \\ N(x_i'\beta,\lambda_i)I(z_i \le 0) \sim \text{otherwise.} \end{cases}$$

The full conditional distribution of λ_i does not have a standard form, but sampling from this distribution is easy using rejection sampling. In Bayesian analysis of the binary logit model, simulations from the full conditional distributions $p(\beta/z,\lambda)$, $p(z/\beta,y)$ and $p(\lambda/z,\beta)$ respectively are implemented using Gibbs sampling. The speed of the simulation in the logit model is slower than that in the probit model because in the logit model it is necessary to simulate from λ in addition to z and β, and the posterior variance-covariance matrix V in Equation (12) changes for each update of λ [4].

4. Illustration

To illustrate the binary logit and probit models a dataset is taken from the Econometric Toolbox in LeSage [5] that contains a binary dependent variable indicating improvement in student grades after exposure to a new teaching method for economics. The explanatory variables are grade point average, a pre-test for understanding of college-level economics (TUCE) and a binary variable indicator showing whether the student was exposed to the new teaching method. For this data set, three estimation methods (OLSs, ML and Bayesian) are used. The MATLAB program is used to implement these methods.

The MATLAB functions **logit** and **probit** are used to find the OLSs and ML estimates for the binary logit and probit models, whereas the **probit_g** function is used to find the Bayesian estimates. If the user assigns a large hyperparameter value $r = 100$ and a diffuse prior for β, the parameter estimates are close to those for the traditional probit model. On the other hand, setting $r = 7$ and assigning a diffuse prior for β should produce estimates close to those from a traditional logit regression.

In the resulting printouts, the **logit** and **probit** functions display the usual coefficient estimatesi t-statistics, marginal probabilities and measures of fit proposed by McFadden [**?**] and Estrella [**?**]. The **probit_ g** function displays posterior estimates of the parameters and iteration numbers in addition to the measures displayed by **logit** and **probit** functions. Besides, the **coda** function displays the convergence diagnostics of the Bayesian estimates, which were mentioned in Section 2. In Table 1, some of these measures for OLSs estimates, MLEs and Bayesian estimates are given for the models concerned.

Table 1. Measures of fit (R^2) proposed by McFadden and the significance level of each parameter estimates for OLSs, ML and Gibbs estimates.

	OLS	ML		Bayesian	
		Logit $R^2 = 0.3740$	Probit $R^2 = 0.3775$	Logit $R^2 = 0.3700$	Probit $R^2 = 0.3752$
Variables	*p-level*	*p-level*	*p-level*	*p-level*	*p-level*
const	0.007929	0.013382	0.006656	0.000000	0.000500
psi	0.011088	0.033617	0.023445	0.011500	0.005000
tuce	0594361	0.506944	0.542463	0.273500	0.261500
gpa	0.007841	0.033376	0.026459	0.010000	0.003000

* : In the Ordinary Least Squares method, the measure of fit value cannot be found by McFadden's approach.

Since OLSs estimates for binary regression models don't ensure the model assumptions, logit and probit regression models based on these estimates cannot be used to make reliable interpretations. The R^2 value for the OLS method cannot be compared to those of the other estimation methods since it is not based on McFadden's approach.

All models give similar results for this data and all models are significant. In the ML and Bayesian methods, the logit and probit models give very similar results. Generally, when the dependent variable is binary, logit and probit model results are similar. McFadden's R^2 value tends to be smaller than the classical R^2 value, and values of 0.2 and 0.4 are considered highly satisfactory. In the models which we examine, McFadden's R^2 values are close to 0.4 so that all models are highly satisfactory. These values are a bit larger in the probit models than those in the logit models. For this reason it can be said that the probit model is more appropriate.

The **coda** function calculates some convergence diagnostics for Bayesian estimates. Autocorrelations and Raftery & Lewis convergence diagnostics are given in Tables 2a and 2b, 3a and 3b (Based on a sample size of 2000).

Table 2a. Autocorrelations within each parameter chain in the logit model

Variable	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**
const	0.315	0.150	0.065	0.047
psi	0.311	0.110	0.046	-0.011
tuce	0.236	0.094	0.076	0.056
gpa	0.277	0.130	0.002	0.032

Table 2b. Raftery-Lewis Diagnostics for each parameter chain in the logit model

Variable	Thin	Burn	Total(N)	(Nmin)	I-stat
const	2	9	2311	937	2.466
psi	2	9	2311	937	2.466
tuce	2	9	2311	937	2.466
gpa	2	9	2311	937	2.466

Table 3a. Autocorrelations within each parameter chain in the probit model

Variable	Lag 1	Lag 5	Lag 10	Lag 50
const	0.070	0.031	0.216	0.015
psi	0.042	0.011	0.153	0.001
tuce	-0.075	-0.096	0.152	-0.044
gpa	0.042	0.023	0.170	0.006

Table 3b. Raftery-Lewis Diagnostics for each parameter chain in the probit model

Variable	Thin	Burn	Total(N)	(Nmin)	I-stat
const	2	12	2350	937	2.508
psi	2	12	2350	937	2.508
tuce	2	12	2350	937	2.508
gpa	2	12	2350	937	2.508

The autocorrelations are not very significant for both models except the the **lag 1** autocorrelations for the logit model. After **lag 1** the autocorrelations become very small for the logit model which means that the chains converge rapidly. The **coda** output reports **Nmin** - the minimum number of iterations that would be needed to estimate the specified quantile to the desired precision if the samples in the chain were independent; **Total(N)** - the total number of iterations that should be run for each variable; **Burn** - the number of initial iterations to discard as the "burn-in"; and **Thin** - the thinning interval to be used.

The final column in the **coda** output reports $\mathbf{I} = \mathbf{N/Nmin}$. This measures the increase in the number of iterations needed to reach convergence due to dependence among the samples in the chain. Values of **I** much greater than 1 indicate high correlations and probable convergence failure. Raftery and Lewis suggest that $\mathbf{I} > 5$ often indicates convergence problems with the sampler. The thinning interval for these models is 2. The number of iterations to be discarded are 9 and 12 for the logit and probit models respectively. The minimum number of iterations for both models are the same, but the total number of iterations for the probit model is a bit larger than for the logit model. Since **I-stat** values are considerably less than 5 for the logit and probit models, there are no convergence problems with the sampler.

The **coda** function also reports estimates of the NSE and RNE based on 4%, 8% and 15% tapering or truncation of the periodgram window. Differences among these estimates reflect autocorrelation in the draws. For this data set it can be said that there is no autocorrelation in the draws for both models because these estimates are not very different. The second set of diagnostics suggested by Geweke determine whether the chains reach an equilibrium state by comparing the means of the first 20% and the last 50% of the sample draws. From the **coda** output it can be said that the chain of draws from the Gibbs sampler has reached an equilibrium state.

The convergence behaviour of the parameter chains can also be seen from their plots. The plots of β_1 are displayed in Figures 1a and 1b.

Figure 1a. The plot of β_1 in the logit model

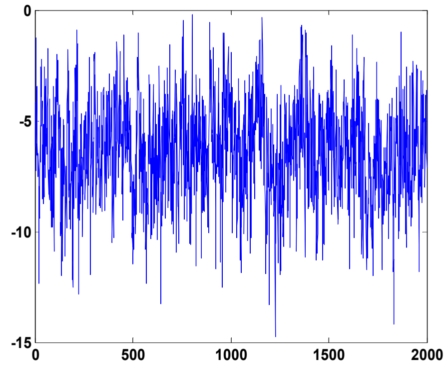
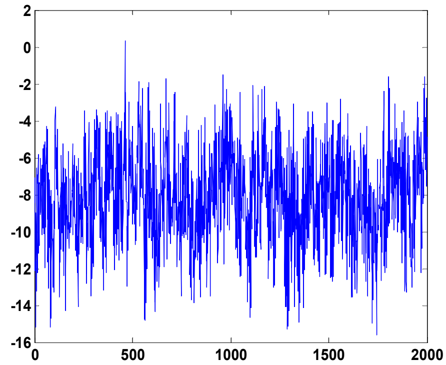


Figure 1b. The plot of β_1 in the probit model



From Figures 1a and 1b, it can be said that the convergence behaviour of β_1 in the logit and probit model is very similar. For both models the chains belonging to β_1 have good mixing. The plots of β_2, β_3, and β_4 can be derived in the same way, and these plots are also seen to be very similar in the logit and probit models.

The graphs of the posterior density functions (pdf) of the parameters can also be drawn. The graphs for β_1 are shown in Figures 2a and 2b, respectively, for the logit and probit models.

Figure 2a. The pdf of β_1 for the logit model

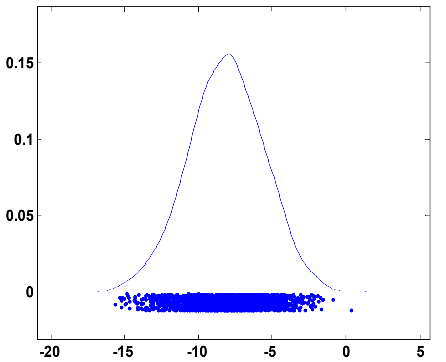
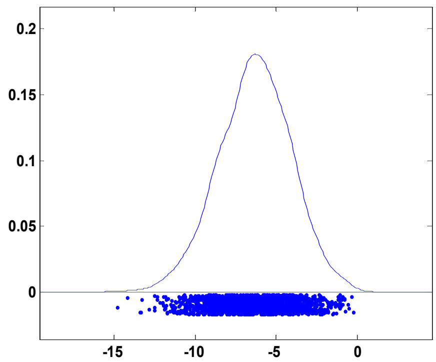


Figure 2b. The pdf of β_1 for the probit model



The pdf graphs of β_1 for the logit and probit models are very similar. The pdf graphs of the other parameters can also be drawn in the same way, and these graphs are also found to be very similar in the logit and probit models.

5. Conclusion

The aim of this article is to use the Bayesian method proposed by Albert and Chib [1] to analyze binary logit and probit models, and to compare this method with other classical methods. The most important point of this Bayesian approach is that the binary probit model is obtained from the normal linear regression model by introducing latent variables into the model. This approach has many advantages. Firstly, in small samples this method provides accurate inferences about the model, whereas the classical inferences in small samples are not accurate. Secondly, in this Bayesian method Gibbs sampling is used to calculate the posterior distributions of the model parameters by simulating from standard distributions. Thus, this method can be implemented easily in many computer programs. Lastly, probit model can be easily analyzed by using a suitable mixture of normal distributions to model the latent data.

One caution in the use of Gibbs sampling is that extra randomness is introduced into the estimation procedure by simulation, and it is important to understand when a particular simulation process has converged. Raftery and Lewis's and Geweke's methods for the diagnosis of convergence have been discussed here, and in the example in Section 5 it has been seen that both the logit and probit models converge rapidly.

References

[1] Albert, J.H. and Chib, S. *Bayesian Analysis of Binary and Polychotomous Response Data*, Journal of the American Statistical Association **422**, 669–679, 1993.

[2] Gelfand, A.E. and Smith, A.F.M. *Sampling-Based Approaches to Calculating Marginal Densities*, Journal of the American Statistical Association **85**, 398–409, 1990.

[3] Geweke, J. *Evaluating the Accuracy of Sampling Based Approaches to the calculation of Posterior Moments*, in: (Bayesian Statistics 4, Oxford University Press, 1992), 169-194.

[4] Holmes, C.C. and Held, L. *Bayesian Auxiliary Variable Models for Binary and Multinomial Regression*, Bayesian Analysis **1**, 145–168, 2006.

[5] Lesage, J.P. *Applied Econometrics using MATLAB* (e-book: `http://www.spatialeconometrics.com`, 1999).

[6] Raftery, A.E. and Lewis, S. *How many iterations in the Gibbs Sampler?*, in: (Bayesian Statistics 4, Oxford University Press, 1992), 763–773.

[7] Tanner, M.A. *Tools for Statistical Inference* (Springer-Verlag, NewYork, 1993).

[8] Tanner, M.A. and Wong, W.H. *The Calculation of Posterior Distributions by Data Augmentation*, Journal of the American Statistical Association **82**, 528–550, 1987.

[9] Walsh, B. *Markov Chain Monte Carlo and Gibbs Sampling*, Lecture Notes for EBB, `http://www.nitro.biosci.arizona.edu/courses/EEB596/handouts/Gibbs.pdf`, 2002.