THE ANALYSIS OF BAYESIAN PROBIT REGRESSION OF BINARY AND POLYCHOTOMOUS RESPONSE DATA

Shadi Nasrollahzadeh

Department of Mathematics, Teachers' Training Faculty
Tehran South Branch, Islamic Azad University
Tehran, Iran
sh_nasrollah@yahoo.com

Abstract   The goal of this study is to introduce a statistical method regarding the analysis of specific latent data for regression analysis of the discrete data and to build a relation between a probit regression model (related to the discrete response) and normal linear regression model (related to the latent data of continuous response). This method provides precise inferences on binary and multinomial models which particularly in the case of small samples, has preference to maximum likelihood methods. The probit regression model for binary outcomes can be easily and precisely explained using different normal distributions for latent data modeling. Applying this approach and using Gibbs sampler method needs simulation of standard distributions such as multivariate normal distribution. Therefore, it can be easily implemented by many softwares and it provides a general method for analyzing binary (or polychotomous) response regression models.

<div dir="rtl">

چکیده   هدف از این مقاله معرفی روش آماری مربوط به استفاده از داده‌های نهانی خاص برای تحلیل رگرسیونی داده‌های گسسته و برقراری ارتباط بین مدل رگرسیون پروبیت (مربوط به پاسخ گسسته) و مدل رگرسیون خطی نرمال (مربوط به داده‌های نهانی پاسخ پیوسته) است. این روش امکان می‌دهد که استنباط های دقیقی برای مدل های رگرسیون دوحالتی و چندحالتی انجام دهیم که به ویژه در نمونه‌های کوچک، به روش حداکثر درست نمایی برتری دارد. مدل رگرسیون پروبیت برای نتایج دوحالتی را با استفاده از توزیع های نرمال مختلف برای مدل بندی داده‌های نهانی به آسانی و بادقت می‌توان توجیه کرد. بکارگیری این ابزار و استفاده از روش نمونه‌گیر گیبز به شبیه سازی توزیع های استاندارد مانند توزیع نرمال چند متغیره نیاز دارد که اجرای آن با استفاده از بسیاری از نرم‌افزارهای کامپیوتری آسان است و روشی عام را برای تحلیل و بررسی مدل های رگرسیون پاسخ دوحالتی (و چند حالتی) فراهم می‌سازد.

</div>

1. INTRODUCTION

A vast literature in statistics, biometrics, and econometrics is concerned with the analysis of binary and polychotomous response data. For example in pharmacognosia tests, after using different viscosity of special poison on limited time duration, the researchers count the dead insects and they often want to find the best relation between the ratio of death and viscosity of poison.

In statistical researches when independent variables and dependent variables (response) are continuous, regression methods are used. If independent variables are discrete and response variables are continuous the analysis of variance

methods, and if some of the independent variables are discrete and the others, continuous, the analysis of covariance methods will be used.

But, in cases where the response variable is discrete (regardless of the type of independent variables), nonlinear methods are used. These are the main parts of statistical researches, especially those in the biometry field. As, these investigations are very important in applied sciences, their analysis methods will be discussed.

When the response variable is discrete and polychotomous (and there is one or several polynomial distributions), it is not simple to evaluate the likelihood function for estimating polynomial probabilities with common methods

(classic, Bayesian and maximum likelihood), because of their special defection [1].

The regression relation of dependent variable Y in terms of one or several independent variables X is E(Y|X) = XB. In the case of binary discrete response variable instead of data, different probabilities are considered proper non linear model for regression relation is used. These nonlinear models are shown as: E(Y|X) = P(Y=1) = P = H(X)

H(.) is a certain cumulative distribution function called link function and it links conditional expectations of variable Y to the independent variable(s) X. If in nonlinear regression models for discrete data, the variation of Y in terms of X is approximately logarithmic, then H is considered as standard normal cumulative distribution function (Φ) and the probit model is obtained, whereas the logit model is obtained if H is the logistic cdf [2]. If N is observed independently of the binary random variables y_i (i=1,…N) with success probabilities P_i (P_i parameters are related to the discrete or continuous auxiliary variables by link function), the binary probit regression model is defined as follows

$$\Rightarrow y_i = \Phi(x_i'\beta) + u_i$$

$$E(y_i|x_i) = P_i = \Phi(x_i'\beta) \quad i = 1,...,N$$

Where

$$u_i = \begin{cases} 1 - \Phi(x_i'\beta) & \text{if} \quad y_i = 1 \\ -\Phi(x_i'\beta) & \text{if} \quad y_i = 0 \end{cases}$$

So, the residuals do not have sufficient information to define outliers. While in (normal) Bayesian regression models, the residuals have a continuous distribution on an interval, and they are more efficient to define the outliners;

$$Z_i = x_i'\beta + \varepsilon_i \qquad i = 1,..,N$$

$$Z_i \sim N(x_i'\beta, \sigma_i^2) \quad , \quad \varepsilon_i \sim N(o, \sigma_i^2)$$

Consequently, a relation is made between the binary probit regression models and (normal) Bayesian regression models. In this paper, a method is introduced to estimate regression coefficients vector (β), with iterated sampling of (standard or nonstandard) conditional distributions, through the Bayesian methods under normal linear regression model. The Gibbs sampler is a modification of the Metropolis algorithm [3]. The Metropolis algorithm was developed to investigate the equilibrium properties of large systems of particles such as molecules in a gas.

Hasting [4] suggests Markov's chain methods of sampling that generalize the Metropolis algorithm. Li [5,6] appears to have independently developed the Gibbs sampler in the context of multiple imputation. In this paper the Gibbs sampler algorithm is used with a focus on its implementation in the binary and polychotomous reponse models [7]. This approach is very similar to the data augmentation formework used in cencored regression models [8].

If in Bayesian analysis for attaining effect estimation, minimizing the $E(\hat{\beta}-\beta)^2$ is considered, the final estimation would be the posterior distribution mean. Then, the polynomial probabilities are estimated by the relation between these probabilities and the linear combination of regression coefficients.

As an application, the analysis of the Bayesian regression method regarding effects of fourteen different environments on relative frequency of five kinds of special insects, based on the results of a real experience, is mentioned in this paper.

2. BAYESIAN PROBIT REGRESSION

The probit regression model for binary outcomes is seen to have an underlying normal regression structure on latent continuous data [9]. Values of the latent data can be simulated from suitable truncated normal distribution. If the latent data are known, then the posterior distribution of parameters can be computed using standard results for normal linear models. Draws from these posteriors are used to sample new latent data and the process is iterated with the Gibbs sampling.

This data augmentation approach provides a general framework for analyzing binary regression models. Under the proposed framework, the class of probit regression models can be enlarged by

using mixtures of normal distributions to model the latent data [10] (the mixtures of normal distributions are normal distributions that their means and variances, or both are definite function of random variables with specific distributions). In this normal mixture class, one can investigate the sensitivity of the parameter estimates to the choice of "link function". The method can also be generalized to multinomial response models.

3. THE GIBBS SAMPLER

An algorithm for extracting the marginal distributions from these full conditional distributions was formally introduced by Geman and Geman (1984) [11,12].

The Gibbs sampler was developed and has been mainly applied in the context of complex stochastic models involving very large numbers of variables [13]. The Gibbs sampling is a Markovian updating scheme that proceeds as follows:

Given an arbitrary starting set of values

$$U_1^{(0)},...,U_k^{(0)}$$

Observation $U_1^{(1)}$ from

$\pi(U_1 | U_2^{(0)},...,U_K^{(0)})$ observation $U_2^{(1)}$ from

$\pi(U_2 | U_1^{(1)}, U_3^{(0)},...,U_K^{(0)})$ ...

and

observation $U_k^{(1)}$ from $\pi(U_K | U_1^{(1)}, U_2^{(1)},...,U_{K-1}^{(0)})$ will be produced.

3.1. Convergence (Densities are denoted generically by brackets).

$$(U_1^{(i)},...,U_k^{(i)}) \xrightarrow{d} [U_1,...,U_K]$$

and hence for each S if $i \to \infty$

$$U_S^{(i)} \xrightarrow{d} U_S \sim [U_S]$$

The Gibbs sampling through m replications of the aforementioned i iterations produces m iid k tuples $(U_{1j}^{(i)}, U_{2j}^{(i)},..., U_{kj}^{(i)})$ (j=1,2,…,m), with the proposed density estimate for $[U_S]$ having form

$$[\hat{U}_S]_i = \frac{1}{m} \sum_{j=1}^{m} [U_S | U_t = U_{(tj)}^i ; t \neq s]$$

4. DATA AUGMENTATION AND GIBBS SAMPLING FOR BINARY DATA

Introduce N latent variables $Z_1,…, Z_N$ where the Z_i are independent N(x'$_i\beta$, 1) and [14] define $Y_i = 1$ if $Z_i > o$ and $Y_i = 0$ otherwise. Then, $Y_i \overset{id}{\sim} Ber(P_i)$

$$P_i = P(Y_i = 1) = P(Z_i > 0) = P(U > -x_i'\beta) =$$

$$P(U < x_i'\beta) = \Phi(x_i'\beta)$$

So, the Z_i, given the data y_i follows a truncated normal distribution. By introducing the Z_i's in to the model, the probit regression model on the Bernoulli observations Y is seen to have an underlying normal regression on Z_i $Z_i = x_i'\beta + \varepsilon_i ; \varepsilon_i \sim N(0,1)$.

As an example [15], consider binary probit regression on target variables $y_n \in \{0,1\}$, the probit likelihood for the nth data sample taking unit value (y_n=1) is $P(y_n=1|x_n,\beta) = \Phi(x_n'\beta)$. Now, this can be obtained by the following marginalization

$$\int p(y_n = 1, Z_n | x_n, \beta)dZ_n =$$
$$\int P(y_n = 1 | Z_n)p(Z_n | x_n, \beta)dz_n$$

and as by definition $p(y_n = 1 | Z_n) = \delta(Z_n > 0)$ then it can be seen that the required marginal is simply the normalizing constant of a left truncated univariate Gaussian so that

$$P(y_n = 1 | x_n, \beta) =)\int \delta(Z_n > 0)$$
$$N_{Z_n}(x_n'\beta,1)dZ_n = \Phi(x_n'\beta).$$

The key observation here is that working with the

joint distribution

$$P(y_n = 1, Z_n \mid x_n, \beta) = \delta(Z_n > 0) N_{Z_n}(x_n'\beta, 1)$$

provides a straightforward means of Gibbs sampling from the parameter posterior which would not be the case if the marginal term, $\Phi(x_n'\beta)$, was employed in defining the joint distribution over data and parameters.

This data augmentation strategy can be adopted in developing efficient methods to obtain binary and multi-class Gaussian process (GP) classifiers [16].

By computation of the marginal posterior distribution of β using the Gibbs sampling algorithm requires only the posterior distribution of β conditional on Z and the posterior distribution of Z conditional on β, and these fully conditional distributions are of standard forms

$$\pi(\beta \mid Y, Z) = C\pi(\beta) \prod_{i=1}^{N} \Phi(Z_i ; x_i'\beta, 1)$$

If a priori the distribution of β is diffuse, then

$$\beta \mid Y, Z \sim N_K((X'X)^{-1}(X'Z), (X'X)^{-1})$$

If β is assigned the proper conjugate $N(\beta^*, B^*)$ prior, then the posterior distribution of β given Z is $N_k(\widetilde{\beta}, \widetilde{B})$.

where

$$\widetilde{\beta} = (B^{*-1} + X'X)^{-1}(B^*\beta^* + X'Z)$$

$$\widetilde{B} = (B^{*-1} + X'X)^{-1}$$

The posterior of Z, conditional on β, also has a simple form truncated normal distribution as follow:

$Z_i \mid y, \beta$ distributed $N(x_i'\beta, 1)$

Truncated at the left by o if $y_i = 1$

$Z_i \mid y, \beta$ distributed $N(x_i'\beta, 1)$

Truncated at the right by o if $y_i = o$

Given a previous value of β, one cycle of the Gibbs algorithm would produce Z and β from the mentioned distributions. The starting value of β, $\beta^{(o)}$ may be taken to be the maximum likelihood (ML) estimate, or the least squares (LS) estimate $(x'x)^{-1}x'y$.

Since the posterior distribution of β given Z is multivariate normal, it is possible to generalize this model by applying suitable mixtures of normal distributions.

For example, one can generalize the probit link by choosing the link cdf H to be the family of t distributions. Let Z_i be independently distributed from t distributions with locations $x_i'\beta$, scale parameter 1 and degree of freedom r. the additional random variable λ_i is introduced and distributed Gamma (r/2, 2/r). $Z_i \mid \lambda_i$ is distributed N ($x_i'\beta, \lambda_i^{-1}$).

Suppose a uniform prior is chosen for β and $\lambda = (\lambda_1, \ldots, \lambda_N)$ be the vector of scale parameters. The fully conditional distributions of β, Z, λ and r are given below:

$$\beta \mid y, Z, \lambda, r \sim N_k(\hat{\beta}_{Z,\lambda}, (X'WX)^{-1})$$

Where

$$\hat{\beta}_{Z,\lambda} = (X'WX)^{-1}(X'W^{-1}Z), \quad W = \text{diag}(\lambda_i)$$

$$Z_i \mid y, \beta, \lambda, r \sim N(x_i'\beta, \lambda_i^{-1})$$

Truncated at the left by o if $y_i = 1$

$$Z_i \mid y, \beta, \lambda, r \sim N(x_i'\beta, \lambda_i^{-1})$$

Truncated at the right by o if $y_i = o$

$$\lambda_i \mid y, Z, \beta, r \sim \text{Gamma}\left(\frac{r+1}{2}, \frac{2}{r + (Z_i - x_i'\beta)^2}\right) \quad r \mid y, Z, \beta,$$

λ is distributed according to the pdf proportional to

Where

$$C(r) = \left[\pi(\frac{r}{2})(\frac{r}{2})^{\frac{r}{2}}\right]^{-1}$$

and $\pi(r)$ is the prior on r.

It started with $\hat{\beta} = (X'X)^{-1}X'y$ (LSE), set $\lambda_i = 1$ for all i, and cycle through the conditional distributions Z, β, λ, r, in that order.

For making inferences about the regression vector β and the probabilities

$\left\{ P_k \right\}$, The posterior for β is approximated

by $\hat{\pi}(\beta) \approx \dfrac{1}{m} \sum\limits_{i=1}^{m} \pi(\beta | Z^{(i)}, \lambda^{(i)})$.

To obtain a posterior density estimate for P_k, it is known that

$$P_k = \Phi(\lambda_k^{\frac{1}{2}} x_k' \beta).$$

Then by a transformation, the density estimate of the probability is given by

$$\hat{\pi}(P_k) = \frac{1}{m} \sum\limits_{i=1}^{m} \phi(\Phi^{-1}(P_k); \mu, \sigma^2) / \phi(\Phi^{-1}(P_k); o, 1)$$

Where

$\phi(\ ; \mu, \sigma^2)$ is the $N(\mu, \sigma^2)$ pdf, $\mu = \sqrt{\lambda_k^{(i)}}\ x_k'\ \hat{\beta}_{Z,\lambda}$

and

$$\sigma^2 = \lambda_k^{(i)} x_k' (X'WX)^{-1} x_k$$

5. GENERALIZATION TO A MULTINOMIAL RESPONSE

The data with multinomial distribution is called polychotomous response data. Each y_i can be considered as a special category. Then there are two different situations, multinomial probit model with

5.1. Orderd Categories
5.2. Unordered Categories

5.1. Ordered Categories Suppose that $y_1, \ldots, y_N$ are observed, where y_i takes one of J ordered categories, $1, \ldots, J$.

and

$$P_{ij} = p(Y_i = j) \qquad\qquad i = 1, 2, \ldots, N$$
$$j = 1, 2, \ldots, J$$

The cumulative probabilities are defined

$$\xi_{ij} = \sum\limits_{k=1}^{j} p_{ik} \qquad\qquad j = 1, 2, \ldots, J-1$$

The latent continuous random variable Z_i distributed $N(x_i'\beta, 1)$

and y_i is observed Where $y_i = j$ if $\gamma_{j-1} < Z_i \le \gamma_j$ ($\gamma_J = \infty$, $\gamma_0 = -\infty$). Then the regression model for the $\left\{ P_{ij} \right\}$ is given by $\xi_{ij} = \Phi(\gamma_j - x_i'\beta)$.

If a diffuse prior for (β,γ), is assigned then $Z_i | \beta, \gamma, Y_i = j \sim N(x_i'\beta, 1)$ truncated at the left (right) by γ_{j-1} (γ_j). Finally, γ_j given Z, y, β can be seen to be uniform on the specified interval. To implement the Gibbs sampler, start with (β,γ) set equal to the MLE and simulate form mentioned distributions. After getting the result of Bayes estimate of β (posterior distribution mean) and using normal regression model

$$\xi_{ij} = \Phi(\gamma_j - x_i'\beta)$$

polynomial probabilities $\left\{ P_{ij} \right\}$ can be calculated.

5.2. Unordered Categories Probit model for polynomial data is considered as follows

$$P_{ij} = \Phi(x_i'\beta) \qquad\qquad i = 1, 2, \ldots, N$$
$$j = 1, 2, \ldots, J$$

i is the index of experimental units and j is the index of categories. In unit i, one of the possible J results with probabilities of $P_{i1}, \ldots, P_{iJ}$ observed. Latent data $Z_i = (Z_{i1}, \ldots, Z_{iJ})$ was introduced and normal regression model

$$i = 1,2,\ldots,N$$
$$j = 1,2,\ldots,J$$

$$Z_{ij} = x'_{ij}\beta + \varepsilon_{ij}$$

$$\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{ij})'$$

Where

$N_J(0,\Sigma)$ is distributed

Σ is a JxJ matrix that is parameterized in terms of parameter vector θ of dimension not exceeding $\dfrac{J(J-1)}{2}$.

Category j is observed if $Z_{ij} > Z_{ik}$ and the multinomial probabilities are given by $P_{ij} = P(Z_{ij} > Z_{ik} \quad ; \quad \forall k \neq j) = P(x'_{ij}\beta + \varepsilon_{ij} >$

So, computation of these probabilities entails calculation of multiple integrals of the multivariate normal density; thus maximum likelihood estimation is very difficult to perform for Large J.

By the following the Gibbs sampling approach, the computation of the multinomial probabilities can be avoided. The vector of observed categories is denoted as $Y = (y_1,\ldots,y_N)$ where $y_i \in \{1,\ldots,J\}$ and $x_i = (x_{i1},\ldots,x_{iJ})'$, the preceding model can be rewritten as

$Z = X\beta + \varepsilon$ where $\varepsilon = (\varepsilon'_1,\ldots,\ldots,\varepsilon'_N)'$ is distributed $N_{NJ}(o, \Omega = I_N \otimes \Sigma)$.

The operator $\otimes$ is the direct multiplication of two matrixes and it is defined as follow

$$C = A_{P\times P} \otimes B_{n\times n} = \begin{bmatrix} a_{11}B\ldots a_{1p}B \\ \vdots \\ a_{p1}B\ldots a_{pp}B \end{bmatrix}$$

To implement the Gibbs sampler, samples are required to form the following conditional distributions:

$$\beta \mid Y, Z_1,\ldots,Z_N, \theta$$
$$Z_1,\ldots,Z_N \mid Y, \beta, \theta$$
$$\theta \mid Y, Z_1,\ldots,Z_N, \beta$$

If a diffuse prior is placed on β, then standard multivariate normal theory yields that

$$\beta \mid Z_1,\ldots,Z_N, Y, \theta \sim N_K(\hat{\beta}_Z, (X'\Omega^{-1}X)^{-1}) \quad \text{where}$$
$$\hat{\beta}_Z = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Z)$$

The vector $\{Z_i\}$ given Y, β, θ is an independent collection with $T^2 = n\hat{\beta}'_1 S_1^{-1} \hat{\beta}_1 = 324.16$ such that the y_i th component of Z_i is the maximum. Using a prior $\pi(\theta)$ on θ, the density of this distribution is proportional to

$$f(\theta) = \pi(\theta)|\Omega(\theta)|^{-\frac{1}{2}} \exp$$
$$\left\{ \frac{-1}{2}(Z - X\beta)'\Omega^{-1}(\theta)(Z - X\beta) \right\}$$

This distribution is not a familiar parametric family and is relatively difficult to simulate. However, in Bayesian methods, the conditional distribution mode can be introduced as an estimation, after calculating the posterior distribution. The considered mode can be calculated by using the methods of differentiating of matrix functions [17] as follows

$$\partial f(\theta) \Big/ \partial\theta = \frac{-1}{2}|\Omega(\theta)|^{\frac{-3}{2}}|\Omega(\theta)|\Omega^{-1}(\theta)\exp$$
$$\left\{ \frac{-1}{2}(Z - X\beta)'\Omega^{-1}(\theta)(Z - X\beta) \right\}$$
$$+ \frac{1}{2}\Omega^{-1}(\theta)(Z - X\beta)(Z - X\beta)'\Omega^{-1}(\theta)$$

$$\exp\left\{ \frac{-1}{2}(Z - X\beta)'\Omega^{-1}(\theta)(Z - X\beta) \right\}|\Omega(\theta)|^{\frac{-1}{2}} = o$$

$$\Rightarrow \frac{-1}{2}|\Omega(\theta)|^{\frac{-1}{2}}\Omega^{-1}(\theta)\exp$$
$$\left\{ \frac{-1}{2}(Z - X\beta)'\Omega^{-1}(\theta)(Z - X\beta) \right\}$$

$$(I - (Z - X\beta)(Z - X\beta)'\Omega^{-1}(\theta) = o$$

$$\Rightarrow I - (Z - X\beta)(Z - X\beta)'\Omega^{-1}(\theta) = o$$

$$\Rightarrow \Omega(\theta) - (Z - X\beta)(Z - X\beta)' = o$$
$$\Rightarrow \hat{\Omega}(\theta) = (Z - X\beta)(Z - X\beta)'$$

So, one way for estimating θ is to calculate $R_i = Z_i - X_i\hat{\beta}$ and $\hat{\Sigma}_i = R_i R'_i$ in each of the implementation of the Gibbs sampler. Then the estimation of matix Σ can be the arithmetic average of $\hat{\Sigma}_i$ for $i = 1,....,N$.

Considering the initial values $\hat{\Sigma}^{(o)} = I$ and $\hat{\beta}^{(o)} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$, the samplings implement of the fully conditionally distributions in the order of Z, β, θ (and calculate the estimation of Σ in each time). After implementing the Gibbs sampler with the different initial points $m = m^*$, which are iterated $t = t^*$ times, the mean and variance of the marginal posterior distributions could be calculated and the average of vectors can be introduced as Bayes estimation [18]. Incidentally, in the most of the practical cases $t \leq 50$ and $m \leq 100$ were proper for convergence.

6. A NUMERICAL EXAMPLE

6.1. Experiment and Design Although it has been well documented that coarse woody debris are important in temperate forest ecosystems, little is known on whether or not these debris have any effect on maintaining biodiversity. Thus, the following experiment was conducted in order to find whether or not decaying logs provide microhabitat for the different species of ground-dwelling arthropods. On the upper coastal plain of Georgia, a twelve month study was undertaken at two adjacent sites on the Savanna River Site [19]. Hundreds of species were collected each month.

The dataset contained the following variables (with their respective meanings).The first factor has two levels:

- MH (Mixed Hardwood)
- PP (Pine)

The second factor has seven levels:
- F (Fake Log)
- CD (Control Dry)

- CW (Control Wet)
- GD (Green Log Dry)
- GW (Green Log Wet)
- DD (Decomposed Log Dry)
- DW (Decomposed Log Wet)

The third factor has six levels:

- COLL (Number of collembola collected)
- COLEOP (Number of coleopterans collected)
- HYMEN (Number of hymenpterans collected)
- SPID (Number of spiders collected)
- ORTHOP (Number of orthopterans collected)
- MISC (Number of miscellaneous ground bugs collected)

Five replications were done for each situation.

6.2. Analysis For analyzing this experiment, fourteen combinations of treatments (2×7) and species of ground-dwelling arthropods were considered as independent variable and the number of species as dependent variable. The dataset of the first month are shown in Table 1.

The five replications of observations in the first cell in Table 1 have been 207, 121, 55, 157, 74, therefore the expected number in this cell is 122.8.

The vector of observations Y is the number of observations collected from treatments.

This vector is $Y = (y'_1,\ldots,y'_{14})'$ where $y = (y_{i1},\ldots y_{i6})'$, and each category has the average of five replications.

Vector Z also has the same form. The linear normal regression model is considered

$$Z_{ij} = x'_{ij}\beta + \varepsilon_{ij} \qquad i = 1,\ldots,14\, j = 1,\ldots,6$$

or in matrix form $Z = X\beta + \varepsilon$ where $\varepsilon = (\varepsilon'_1,\ldots,\varepsilon'_{14})$ is distributed $N_{84}(o,\Omega)$ $(\Omega = I_{14} \otimes \Sigma_6)$ and $\varepsilon_I = (\varepsilon_{i1},\ldots\ldots,\varepsilon_{i6})'$ is distributed $N_6(o,\Sigma)$.

The regression coefficients β has the following form

$\beta = (\beta_o,\beta_1,\ldots,\beta_5,\beta_6,\beta_7,\ldots,\beta_{18})'$
Where β_o is the constant of the model and β_j ($j = 1,\ldots 5$) shows the effect of each species and β_j ($j =$

6, ….., 18) shows the effect of the combination of treatment j on number of species.

The independent variables are discrete, the auxiliary variables X_i must be used. The matrix of X (design matrix) is defined as

$$X_{84 \times 19} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{14} \end{bmatrix}$$

and X_i is a 6×19 matrix with proper form. By using the computer program prepared by the author of this paper for the mentioned method and iterating Gibbs sampler for t = 10, m = 40 the mean and variance-covariance matrix of marginal posterior distribution of β and the mean of posterior distribution as Bayes estimation have been calculated for each month. Incidentally, in addition to Bayes estimations, the calculated maximum likelihood estimations, for comparison, are shown in Table 2. It is well found that the results are reasonable and expected.

Variance-covariance matrix S (19×19) $(\hat{\Sigma})$ also is calculated, but because of its large dimensions it has not been shown here.

TABLE 1. Expected Numbers of Insects in Month One.

Kind of Insect / Treat	Coll	Coleop	Hymen	Spid	Orthop	Misc
PP-F	122.8	5.2	6	3.8	0.8	29.4
PP-CD	80.6	8	3.8	1.4	0.6	18.4
PP-CW	85.2	9.4	4.6	1.6	0.4	10.4
PP-GD	116.2	6	5.2	3.8	0.4	19.6
PP-GW	134.4	6	1.8	2.2	0.6	13.8
PP-DD	114	4.6	8.4	3.6	0.4	27.4
PP-DW	101.2	8.2	11	3.2	0.8	19.6
MH-F	59.8	1.4	8.8	1	0.8	3.6
MH-CD	99.4	3.2	7	2.4	2.4	3.6
MH-CW	96.4	2.4	6.8	1.2	1.2	2.8
MH-GD	108.2	1.4	10.6	1.4	0.8	2.4
MH-GW	101.4	2.2	5	1	1.4	3.6
MH-DD	73	10.2	8.6	1.2	0.8	6
MH-DW	48.2	3.4	6.2	4	1.2	13.2

TABLE 2. Estimations of Regression Coefficients in Month One.

ML Estimations	Bayes Estimations
2.067 e + 000	1.909 e + 000
1.450 e + 001	1.431 e + 001
1.751 e + 000	1.635 e + 000
5.292 e + 000	5.194 e + 000
8.148 e − 001	7.513 e − 001
1.979 e − 001	8.187 e − 002
-2.931 e + 000	-2.776 e + 000
-1.261 e + 000	-1.055 e + 000
-2.523 e + 000	-2.325 e + 000
-1.883 e + 000	-1.401 e + 000
-1.474 e + 000	-1.085 e + 000
-3.204 e + 000	-2.883 e + 000
-2.078 e + 000	-1.907 e + 000
-2.037 e + 000	-1.986 e + 000
-3.434 e + 001	6.425 e − 002
2.452 e + 000	2.585 e + 000
-1.143 e + 000	-1.236 e + 000
1.498 e + 000	-1.231 e + 000
-2.073 e + 000	-1.845 e + 000

6.3. Hypothesis Testing Regarding the results, the following hypothesis testing and inferences can be performed:

6.3.1. For each month, hypothesis testing

$$\begin{cases} H_o : \beta = \underline{o} \\ H_1 : \beta \neq \underline{o} \end{cases}$$

has been considered where H_o is "The kind of insects and different environments have no effect on the frequency of insects".

As the matrix Σ is unknown, matrix S is used. The critical region of H_o is

$$T^2 = n(\hat{\beta} - \beta_o)'S^{-1}(\hat{\beta} - \beta_o) > \frac{(n-1)P}{n-P} F_\alpha(P, n-P)$$

Where

$$n = 40 \quad P = 19 \quad \alpha = 0.05 \quad \beta_o = o.$$

For example the value of test statistic T^2 for the first month is

$$T^2 = n\hat{\beta}'_1 S_1^{-1} \hat{\beta}'_1 = 324.16$$

$$\frac{(n-1)P}{n-P} F_\alpha(P, n-P) = 75.86$$

324.16 > 75.86 so, H_o is rejected.

6.3.2. For each month hypothesis testing

$$\begin{cases} H_o : \beta = \beta_o \\ H_1 : \beta \neq \beta_o \end{cases}$$

Where

$$\beta_o = \begin{bmatrix} o_{6 \times 1} \\ \hat{\beta}_{1 \ 13 \times 1} \end{bmatrix}$$

has been considered that H_o is "The kind of insects has no effect on the frequency of insects".

$$T^2 = n(\hat{\beta} - \beta_o)'S^{-1}(\hat{\beta} - \beta_o) = 11895 > 75.86 \text{ so, } H_o$$

is rejected.

6.3.3. Hypothesis testing for each level of factors;

For example in the fourth month: $\begin{cases} H_o : \beta_7 = o \\ H_1 : \beta_7 \neq o \end{cases}$

where H_o is "The environment MH-F has no effect on frequency of insects" (MH-F is the seventh regression coefficient in vector β).

Test statistic $t_c = \dfrac{\hat{\beta}_7}{S_{\hat{\beta}_7}} = \dfrac{-2.77}{\sqrt{5.17}} = -1.218$ where

is in acceptance region $-t_{0.025}^{(39)} < t < t_{0.025}^{(39)}$ therefore, H_o is accepted.

6.3.4. The hypothesis testing for comparing two different regression factors is considered as:

$$\begin{cases} H_o : \beta_i = \beta_j \\ H_1 : \beta_i \neq \beta_j \end{cases} \forall i, j$$

which with $\Delta = \beta_i - \beta_j$ is converted to equivalent

test $\begin{cases} H_o : \Delta = o \\ H_1 : \Delta \neq o \end{cases}$. Statistic $D = \hat{\beta}_i - \hat{\beta}_j$ is distributed

$N(\Delta, \sigma_i^2 + \sigma_j^2 - 2\sigma_i \sigma_j)$ So $t_c = \dfrac{D}{S_D} \sim t_\alpha^{(n-1)}$

$(S_D^2 = S_i^2 + S_j^2 - 2S_i S_j)$ and the critical region of H_o is $t_c > t_\alpha^{(n-1)}$.

For example in the fifth month we can test

$$\begin{cases} H_o : \beta_{14} = \beta_{16} \\ H_1 : \beta_{14} > \beta_{16} \end{cases}$$

where H_1 is "The PP-CW is more effective than MH-DW on frequency of insects $D = \hat{\beta}_{14} - \hat{\beta}_{16} = 0.612$

$$S_D^2 = 0.0408 \rightarrow t_c = \frac{0.613}{0.0408} = 3.034$$

$$t_{0.05}^{(39)} = 1.68 \rightarrow t_c > t_{0.05}^{(39)}$$

Therefore, H_1 is accepted.

For example in the fifth month these hypothesis testing can be performed for comparing each two desired factors.

7. CONCLUSION

The main novelty of this paper, has been that by introducing latent data into the problem, the probit model on the binary response is connected with the normal linear model on the continuous latent data reponse. A simulation-based approach has been introduced for Computing the exact posterior distribution of regression coefficients.

Applying this approach allows one to perform exact inference for binary regression models; this will likely be preferable to ML methods for small samples and is especially attractive in the multinomial setup, where it can be difficult to evaluate the likelihood function.

By using Gibbs sampling for simulation from standard distributions, one is introducing extra randomness into the estimation procedure and it is important to understand when a particular simulation process has converged.

In the numerical example, the role of coarse woody debris on macroarthropod diversity has been shown. Computer results related to estimated regression coefficients, using the proposed data

augmentation method, also the ML estimations for comparison and some possible hypothesis testing for inference have been illustrated. Augmenting binary and polychotomous response models with Gaussian latent variables enables exact Bayesian analysis via Gibbs sampling from the parameter posterior.

Because the relative simplicity of this simulation method in this application, it is believed that there will be future research into the automation of this algorithm so that it can be incorporated into standard statistical software.

8. NOMENCULATURE

Ber:	Bernoulli Distribution
d:	Distributed;
E:	Expected Value;
F:	Snedecor's F Distribution;
H_0:	Null Hypothesis;
H_1:	Alternative Hypothesis;
id:	Independent Distributed;
iid:	Independent and Identically Distributed;
$N(\mu,\sigma)$:	The Normal Distribution with Mean μ and Standard Deviation σ
P:	Dimension of Vector β;
S:	Covariance Matrix of Sample;
t:	Student's t Distribution;
u:	Residual;
$U_k^{(i)}$:	K 'th Observation of i 'th Iteration;
Φ:	Cumulative Distribution Function of Standard Normal Distribution;
β:	Regression Coefficient Vector;
$\hat{\beta}$:	Estimation of β;
β_0:	Initial Value of β;
θ:	The Unknown Vector Parameter of Dimension not Exceeding $\frac{J(J-1)}{2}$;
$\sum$:	J Dimensional Covariance Matrix of Population in Terms of θ;
ξ:	Cumulative Probability;
α:	Significance Level of a Test;
ε:	Error of Estimation;
$\Omega(\theta)$:	NJ Dimensional Covariance Matrix of Population in Terms of θ;

9. REFERENCES

1. Zellner, A. and Rossi, P. E., "Bayesian Analysis of Dichotomous Quantal Response Models", *Journal of Econometrics*, Vol. 25, (1984), 365-393.
2. Nelder, J. A. and McCullagh, P., "Generalized Linear Models", Chapman and Hall, New York, U. S. A., (1989).
3. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, Vol. 21, (1953), 1087-1091.
4. Hasting, W. k., "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", *Biometrika*, Vol. 57, (1970), 97-109.
5. Li, K. H., "Hypothesis Testing in Multiple Imputation with Emphasis on Mixed-up Frequencies in Contingency Tables", Ph.D. Dissertation, University of Chicago, Chicago, U. S. A., (August 1985).
6. Li, K. H., "Imputation Using Markov Chains", *Journal of Statistical Computation and Simulation*, Vol. 30, (1988), 57-79.
7. Albert, J. H. and Chib, S., "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, Vol. 88, (1993), 669-679.
8. Chib, S., "Bayes Inference in the Tobit Censored Regression Model", *Journal of Econometrics*, Vol. 51, (1992), 79-99.
9. Khajeh Nouri, A., "Advanced Statistics and Biometry", Tehran University Publicatian, Tehran, Iran, Chapter 18, (1969).
10. Wei, G. C. G. and Tanner, M. A., "Posterior Computations for Censored Regression Data", *Journal of the American Statistical Association*, Vol. 85, No. 411, (1990), 829-839.
11. Gelfand, A. E. and Smith, A. F. M., "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, Vol. 85, (1990), 398-401.
12. Geman, S. and Geman, D., "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, (1984), 721-741.
13. Zeger, S. L. and Karim, M. R., "Generalized Linear Models with Random Effects; A Gibbs Sampling Approach", *Journal of the American Statistical Association*, Vol. 86, No. 413, (1991), 79-85.
14. Griffiths, W. E., Hill, R. C. and Pope, P. J., "Small Sample Properties of Probit Model Estimators", *Journal of the American Statistical Association*, Vol. 82, No. 399, (1987), 929-937.
15. Girolami, M. and Rogers, S., "Variational Bayesian Multinomial Probit Regression with Gaussioan Process Priors", *Journal of Neural Computation*, Vol. 18, (2006), 1790-1817.
16. Williams, C. K. I. and Rasmussen, C. E., "Gaussian Processes for Regression", *Advances in Neural Processing Systems*, Vol. 8, (1996), 598-604.
17. Press, S. J., "Applied Multivariate Analysis", Holt, Rinehart and Winston, New York, U. S. A., (1972).

18. Tanner, M. A. and Wong, W. H., "The Calculation of Posterior Distributions by Data Augmentation", *Journal of the American Statistical Association*, Vol. 82, No. 398, (1987), 528-540.

19. Vanwieren, S. and Datta, S., "Factors Influencing the Colonization and Diversity of Ground-Dwelling Macroarthropods at Coarse Woody Debris on the Upper Coastal Plain of Georgia", D. A. Crossley Institute of Ecology, University of Georgia, Georgia, USA, (1994).