Journal of the Korean
Data & Information Science Society
1998, Vol. 9, No. 2, pp. 165 ~ 172

Bayesian Estimation for the Multiple Regression with Censored Data : Mutivariate Normal Error Terms [1]

Yong Hwa Yoon [2]

Abstract

This paper considers a linear regression model with censored data where each error term follows a multivariate normal distribution. In this paper we consider the diffuse prior distribution for parameters of the linear regression model. With censored data we derive the full conditional densities for parameters of a multiple regression model in order to obtain the marginal posterior densities of the relevant parameters through the Gibbs Sampler, which was proposed by Geman and Geman(1984) and utilized by Gelfand and Smith(1990) with statistical viewpoint.

Key Words and Phrases: Multiple Regression, Multivariate Normal Error Terms, Censored Data, Bayesian Estimation, Gibbs Sampler.

1. Introduction

Censored data often result from life testing and reaction time experiments where it is a common practice to terminate observation prior to failure or reaction of all sample specimens, and problems requiring regression analysis of censored data arise frequently in practice. Maddala(1983) showed some examples of censored regression in econometrics field, including a labor-supply model for the reservation wage and the market wage.

Schmee and Hahn(1979) proposed an iterative least squares(ILS) method for parameter estimation in linear models with right-censored normally distributed response variables. They obtained an initial least squares fit by treating the censored values as failure times. Then, based upon this initial fit, the expected failure time for each censored observation is estimated. These estimates are used in order to

[1]This research was supported by the Taegu University Research Grant,1996
[2]Professor, Department of Statistics,Taegu University, Taegu,713-714

obtain an updated least squares fit, and new expected failure times are estimated for the censored values.

Aitkin(1981) proposed an EM algorithm for the point estimation of the censored regression based on Schmee and Hahn(1979)'s result. He considered the multiple regression model with right-censored data. Wei and Tanner(1990) applied the data augmentation algorithm in case of simple regression with right-censored data. The algorithm generates or imputes the latent data from the predictive distribution, conditional on the fact the failure time must be larger than the observed event time. They obtain the joint posterior density function based on the augmented data sets, and update the process until the joint posterior density function seems to be stabilized.

We will derive the full conditional densities for parameters of a multiple regression model with censored data, where each error term follows a multivariate normal distribution, by using the Gibbs Sampler.

2. Model

We consider the following regression model:

$$y = \mathbf{X}\underline{\beta} + \underline{\epsilon} \tag{1}$$

where $\mathbf{X}$ is $n \times k$ matrix with rank k and $x_{i1} = 1$, $i = 1, 2, \cdots, n$, $\underline{\beta} = (\beta_1, \cdots, \beta_k)$, $\underline{\epsilon}$ is $n \times 1$ error vector and $\underline{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, $\mathbf{I}_n = n \times n$ identity matrix. Note that β_1 is an unknown intercept and $\beta_2, \cdots, \beta_k$ are $k-1$ unknown coefficients. Also we may reorder the data so that the first m observations are uncensored and the remaining $n-m$ are censored. Let c_j denote the censoring time for the case j, $j = m+1, \cdots, n$ and Z_j be the unobserved failure time for j. We assume that the unobserved falure time Z_j is greater than c_j.

The likelihood function $p(\mathbf{y}|\underline{\beta}, \sigma^2)$ is given by

$$\begin{aligned} p(\mathbf{y}|\underline{\beta}, \sigma^2) &\propto \frac{1}{\sigma^n} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\underline{\beta})'(\mathbf{y} - \mathbf{X}\underline{\beta})\} \\ &\propto \frac{1}{\sigma^n} \exp\{-\frac{1}{2\sigma^2}[\nu s^2 + (\underline{\beta} - \underline{\widehat{\beta}})'\mathbf{X}'\mathbf{X}(\underline{\beta} - \underline{\widehat{\beta}})]\} \end{aligned} \tag{2}$$

where $\nu = n - k$, $\underline{\widehat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and $\nu s^2 = (\mathbf{y} - \mathbf{X}\underline{\widehat{\beta}})'(\mathbf{y} - \mathbf{X}\underline{\widehat{\beta}})$.

Our diffuse prior pdf for the elements of β and σ^2 is given by

$$p(\underline{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}, -\infty < \beta_i < \infty,\ i = 1, 2, \cdots, k, \qquad 0 < \sigma^2 < \infty. \tag{3}$$

Then the joint posterior pdf for the parameters is

$$\begin{aligned} p(\underline{\beta}, \sigma^2|\mathbf{y}) &\propto \frac{1}{\sigma^{n+2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\underline{\beta})'(\mathbf{y} - \mathbf{X}\underline{\beta})\} \\ &\propto \frac{1}{\sigma^{n+2}} \exp\{-\frac{1}{2\sigma^2}[\nu s^2 + (\underline{\beta} - \underline{\widehat{\beta}})'\mathbf{X}'\mathbf{X}(\underline{\beta} - \underline{\widehat{\beta}})]\}. \end{aligned} \tag{4}$$

3. Bayesian Estimation

To apply the Gibbs sampler to the regression model with censored data, we need the following full conditional distributions for unknown quantities including regression coefficients $\beta_1, \beta_2, \cdots, \beta_k$, common variance σ^2 and the unobserved observations $Z_{m+1}, \cdots, Z_n$.

The required full conditional distributions are

$$\begin{aligned} & p(\beta_i|\beta_{-i}, \sigma^2, \mathbf{y}), \\ & \quad \text{where } \beta_{-i} = (\beta_1, \cdots, \beta_{i-1}, \beta_{i+1}, \cdots, \beta_k),\ i = 1, 2, \cdots, k, \\ & p(\sigma^2|\underline{\beta}, \mathbf{y}), \\ & p(Z_j|\underline{\beta}, \sigma^2, y_1, \cdots, y_m, Z_i, i \neq j),\ j = m+1, \cdots, n, \end{aligned}$$

where Z_j is the unobserved failure time of the the case j. From the joint posterior density $p(\underline{\beta}, \sigma^2|\mathbf{y})$, we obtain

$$p(\underline{\beta}|\sigma^2, \mathbf{y}) \propto \exp\{-\frac{1}{2\sigma^2}(\underline{\beta} - \widehat{\underline{\beta}})'\mathbf{X}'\mathbf{X}(\underline{\beta} - \widehat{\underline{\beta}})\}. \tag{5}$$

That is, the conditional distribution of $\underline{\beta}$ given σ^2 is the multivariate normal distribution $N_k(\widehat{\underline{\beta}}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, where $\widehat{\underline{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

To obtain the full conditional distribution for β_i, we reorder the data so that the i-th column vector $\mathbf{x}_i$ of matrix $\mathbf{X}$ corresponding to β_i is the first column of matrix $\mathbf{X}$. Then the full conditional density of β_i of interest can be obtained as that of β_1 based on the reordered data set as follows. For the convenience, let $\underline{\beta}_2 = \underline{\beta}_{-i}$

$$\begin{aligned} \underline{\beta} &= \begin{bmatrix} \beta_1 \\ \underline{\beta}_2 \end{bmatrix}, \quad \widehat{\underline{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\underline{\beta}}_2 \end{bmatrix}, \\ \mathbf{X}'\mathbf{X} &= \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{x}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} h_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \end{aligned} \tag{6}$$

where $\mathbf{X}_2$ is the $n \times (k-1)$ matrix obtained by removal of $\mathbf{x}_1$. Partitioning $\underline{\beta} - \widehat{\underline{\beta}}$ and $\mathbf{X}'\mathbf{X}$ to correspond to the partitioning of $\underline{\beta}$, we can write $p(\underline{\beta}|\sigma^2, \mathbf{y})$ as the product of two factors.

$$\begin{aligned} & p(\beta_1, \underline{\beta}_2|\sigma^2, \mathbf{y}) \\ & \propto \exp\left\{-\frac{h_{11}}{2\sigma^2}[\beta_1 - \widehat{\beta}_1 + \frac{H_{12}}{h_{11}}(\underline{\beta}_2 - \widehat{\underline{\beta}}_2)]^2\right\} \\ & \times \exp\left\{-\frac{1}{2\sigma^2}(\underline{\beta}_2 - \widehat{\underline{\beta}}_2)'(H_{22} - \frac{H_{21}H_{12}}{h_{11}})(\underline{\beta}_2 - \widehat{\underline{\beta}}_2)\right\} \end{aligned} \tag{7}$$

From (7),the full conditional distribution of β_1 is a normal distribution with mean $\widehat{\beta}_1 - \frac{H_{12}}{h_{11}}(\underline{\beta}_{-1} - \underline{\widehat{\beta}}_{-1})$ and variance $\frac{\sigma^2}{h_{11}}$. That is,

$$p(\beta_1|\underline{\beta}_{-1}, \sigma^2, \mathbf{y}) \sim N(\widehat{\beta}_1 - h_{11}^{-1}H_{12}(\underline{\beta}_{-1} - \underline{\widehat{\beta}}_{-1}, \frac{\sigma^2}{h_{11}}). \tag{8}$$

By exchanging the i-th column vector and the first column of the matrix $\mathbf{X}$, we can sample a value of β_i of interest by using the full conditional distribution for β_1 based on reordered data.

Since $p(\underline{\beta}|\mathbf{y})$ can be expressed as follows :

$$\begin{aligned} p(\underline{\beta}|\mathbf{y}) &= \int p(\underline{\beta}, \sigma^2|\mathbf{y})d\sigma^2 \\ &\propto \{\nu z^2 + (\underline{\beta} - \underline{\widehat{\beta}})'\mathbf{X}'\mathbf{X}(\underline{\beta} - \underline{\widehat{\beta}})\}^{-n/2}, \end{aligned} \tag{9}$$

which is the well known multivariate student t distribution.

Then the full conditional distribution of σ^2 is given by

$$p(\sigma^2|\underline{\beta}, \mathbf{y}) \propto \frac{1}{\sigma^{n+2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - \underline{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{y} - \underline{\beta})\}. \tag{10}$$

Hence the full conditional distribution of σ^2 is that of the random variable $(\mathbf{y} - \mathbf{X}\underline{\beta})'(\mathbf{y} - \mathbf{X}\underline{\beta})\chi_n^{-2}$, where χ_n^{-2} is an inverted chi-square random variable with degrees of freedom n.

The sampling of the latent data Z_j from the predictive distribution $p(Z_j|\underline{\beta}, \sigma^2, Z_i, i \neq j) = p(Z_j|\underline{\beta}, \sigma^2, Z_j > c_j)$, $j = m+1, \cdots, n$, where c_j is the censoring time of the case j, can be done similarly as the simple regression case. Let $u_j = \epsilon_j/\sigma$, then

$$p(u_j|\underline{\beta}, \sigma^2, Z_j > c_j) = \frac{\phi(x)}{1 - \Phi(c_{j0})}, \qquad u_j > c_{j0},\ j = m+1, \cdots, n,$$

where $c_{j0} = \frac{c_j - \mathbf{x}_j'\underline{\beta}}{\sigma}$, $\mathbf{x}_j$ is the j-th row of $\mathbf{X}$ corresponding to the case j. Then $Z_j = \mathbf{x}_i'\underline{\beta} + u_j^*$, where u_j^* is generated from a left-truncated standard normal distribution which is the predictive distribution for Z_j.

By using the Rao-Blackwellized density estimators given by Gelfand and Smith (1990), we obtain the following density estimators $p(\beta_i|\mathbf{d}_0)$ of marginal density for β_i,

$$p(\widehat{\beta_i|\mathbf{d}_0}) = \frac{1}{J}\sum_{j=1}^{J} p(\beta_i|\beta_{-ij}^{(t)}, y_1, \cdots, y_m, z_{m+1,j}^{(t)}, \cdots, z_{n,j}^{(t)}) \tag{11}$$

where $i = 1, 2, \cdots, k$, $\beta^{(t)}_{-ij} = (\beta^{(t)}_{1j}, \cdots, \beta^{(t)}_{(i-1)j}, \beta^{(t)}_{(i+1)j}, \cdots, \beta^{(t)}_{k,j})$, $\mathbf{d}_0 = (y_1, \cdots, y_m, c_{m+1}, \cdots, c_n)$ and $(\beta^{(t)}_{1j}, \cdots, \beta^{(t)}_{kj}, \sigma^{2(t)}_j, z^{(t)}_{m+1,j}, \cdots, z^{(t)}_{n,j})$, $j = 1, 2, \cdots, J$ denotes the j-th Gibbs sequence for the j-th vector of initial values after t iterations.

And the full conditional density for β_i is obtained by exchanging $\mathbf{x}_1$ and $\mathbf{x}_i$ of the previous matrix $\mathbf{X}$ and adjusting the values of $\widehat{\beta}_1$, h_{11}, H_{12} and $(\underline{\beta}_{-1} - \widehat{\underline{\beta}}_{-1})$ as mentioned before.

The density estimator $\widehat{p(\sigma^2|\mathbf{d}_0)}$ of marginal density for σ^2 is given by

$$
\begin{aligned}
\widehat{p(\sigma^2|\mathbf{d}_0)} &= \frac{1}{J}\sum_{j=1}^{J} p(\sigma^2|\underline{\beta}^{(t)}, y_1, \cdots, y_m, z^{(t)}_{m+1,j}, \cdots, z^{(t)}_{n,j}) \qquad (12)\\
&= \frac{1}{J}\sum_{j=1}^{J}[\Gamma(\frac{n}{2})2^{n/2}]^{-1}(\sigma^2)^{-(n/2+1)}\exp\left(-\frac{A^{(t)}(\underline{\beta})}{2\sigma^2}\right)\left[A^{(t)}(\underline{\beta})\right]^{n/2}
\end{aligned}
$$

where

$$
\begin{aligned}
A^{(t)}(\underline{\beta}) &= (\mathbf{y}^{(t)}_{0j} - \mathbf{X}\underline{\beta}^{(t)}_j)'(\mathbf{y}^{(t)}_{0j} - \mathbf{X}\underline{\beta}^{(t)}_j),\\
\mathbf{y}^{(t)\prime}_{0j} &= (y_1, y_2, \cdots, y_m, z^{(t)}_{m+1,j}, \cdots, z^{(t)}_{n,j}),\\
\underline{\beta}^{(t)}_j &= (\beta^{(t)}_{1j}, \beta^{(t)}_{2j}, \cdots, \beta^{(t)}_{kj}), \quad j = 1, 2, \cdots, J.
\end{aligned}
$$

If the number of interations t is large enough, the above density estimators are regarded as the true marginal densities for β_i's and σ^2.

4. Numerical Example

Schmee and Hahn(1979) used the results of temperature accelerated life tests on electrical insulation in 40 motorettes in order to illustrate the iterative least squares method. Ten motorettes were tested at each of four temperatures. Testing was terminated at different times at each temperature, resulting in a total of 17 failed units and 23 unfailed ones. The model used to analyze the data assumes as follows : (i) For any temperature, the distribution of time to failure is lognormal. (ii) The standard deviation σ of the lognormal time to failure distribution is constant. (iii) The mean of the logarithm of the time to failure is a linear function of the reciprocal $x = 1000/(T + 273.2)$ of the absolute temperature T, *i.e.*, $y = \beta_1 + \beta_2 x + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. This is often referred to as the Arrehenius relationship.

To find the effect of censorship, we obtain two sets of data from Schmee and Hahn(1979)'s data. The first set consists of the first five observations at each level of the temperarture, this serves to expose the effect of the censored observations at only one level of x. Table 1 shows the transformed values of 20 observations of which 5 observations at level 2.3629 (the temperature T is 150^o) are censored and the censoring rate is 0.25. Based on this data set , the least squares estimate of (β_1 , β_2) is (-4.531 , 3.433) by treating the censored obserbations as failure times.

Table 1 Transformed Observations Censored at Only One level

2.0276	2.1589	2.2563	2.3629
2.6107	2.6107	3.2465	3.9066*
2.6107	2.6107	3.4428	3.9066*
2.7024	3.1284	3.5371	3.9066*
2.7024	3.1284	3.5492	3.9066*
2.7024	3.1584	3.5775	3.9066*

* denotes the censored values

The second data set consists of the first 8 observations at each level of the temperature. Table 2 shows that the transformed values of 32 observations of which all 8 observations at level 2.3629 (the temperature T is 150^o) are censored and one or three are censored at other levels of temperature. This set serves to expose the effect of the observations which can be censored at every level of x. In this case the censoring rate is $15/32 = 0.47$. Based on this data set, the least squares estimate of (β_1 , β_2) is (-5.148 , 3.818) by treating the censored obserbations as failure times.

Table 2 Transformed Observations Censored at Every Level

2.0276	2.1589	2.2563	2.3629
2.6107	2.6107	3.2465	3.9066*
2.6107	2.6107	3.4428	3.9066*
2.7024	3.1284	3.5371	3.9066*
2.7024	3.1284	3.5492	3.9066*
2.7024	3.1584	3.5775	3.9066*
2.7226*	3.2253*	3.6866	3.9066*
2.7226*	3.2253*	3.7157	3.9066*
2.7226*	3.2253*	3.7362*	3.9066*

* denotes the censored values

Figure 1 shows the marginal posterior distributions $p(\beta_1|Y)$ and $p(\beta_2|Y)$ respectively when n=20, where only the observations in the lowest level of the stress can be censored. The number of sets of initial values is $20,000$.We stop the procedure by using the stopping method proposed by Kim(1994). The procedure is terminated if the sum of maximum differences of the consecutive density estimates over subintervals of the support is less than 0.03 .

The mode of $p(\beta_1|Y)$ is -5.201 with 30 iterations, which is less than the value -5.164 obtained by treating the censored observations as missing ones. The mode of $p(\beta_2|Y)$ is 3.836 with 31 iterations, which is somewhat larger than the value 3.819 obtained by treating the censored observations as missing ones. The point estimates of (β_1, β_2) by the iterative least squares(ILS) method and EM algorithm are (-5.545, 3.999) with 16 iterations and (-5.677, 4.062) with 19 iterations, respectively. The estimate by EM algorithm has the largest absolute values among the three estimates and the estimate by the Gibbs sampler is the smallest estimate in absolute value. Figure 2 shows the marginal posterior distributions $p(\beta_1|Y)$ and $p(\beta_2|Y)$ respectively when n=32, where the observations in every level of the stress can be censored. The mode of $p(\beta_1|Y)$ is -4.931 with 37 iterations, which is larger than the value -4.983 obtained by treating the censored observations as missing ones. The mode of $p(\beta_2|Y)$ is 3.789 with 31 iterations, which is larger than the value 3.762 obtained by treating the censored observations as missing ones. The point estimates of (β_1, β_2) by the iterative least squares(ILS) method and EM algorithm are (-6.257, 4.357) with 23 iterations and (-6.788, 4.605) with 33 iterations, respectively. Similarly in case of n=20, the estimate by EM algorithm has the largest absolute values among the three estimates and the estimate by the Gibbs sampler is the smallest estimate in absolute value.

References

1. Aitkin, M.(1981). A Note on Regression Analysis of Censored Data, *Technometrics*, 23, 161-163.

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

3. Geman, S., and Geman, D.(1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

4. Kim, H. J.(1994). Bayesian Inference on $P(X < Y)$ with the Gibbs Sampler, Ph.D.Thesis, Kyungpook National University.

5. Maddala, G. S.(1983). *Limited-dependent and Qualitative Variables in Econometrics*, New York, Cambridge University Press.

6. Schmee, J., and Hahn, G. J.(1979). A Simple Method for Regression Analysis with Censored Data, *Technometrics*, 21, 417-432.

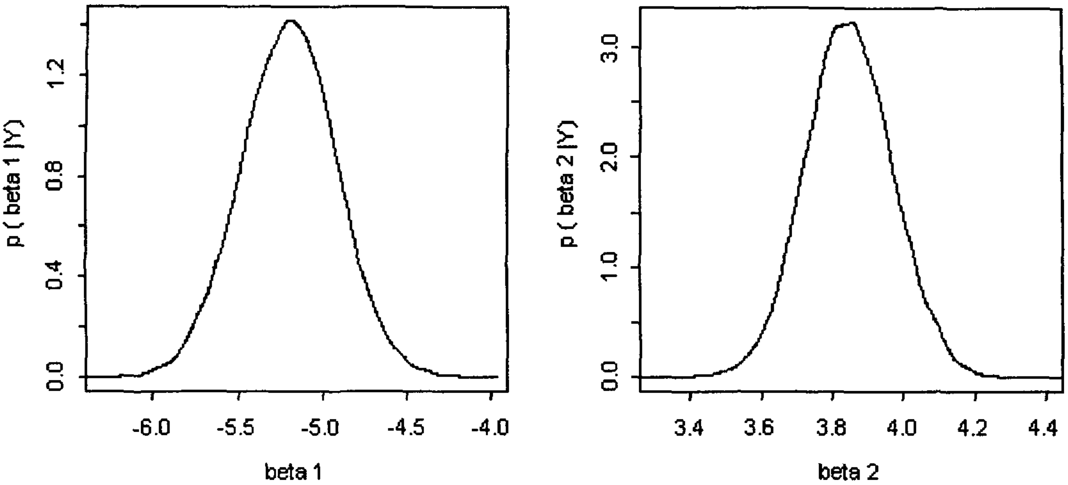

Figure 1 $p(\beta_1|Y)$**(left) and** $p(\beta_2|Y)$**(right) When** n**=20**

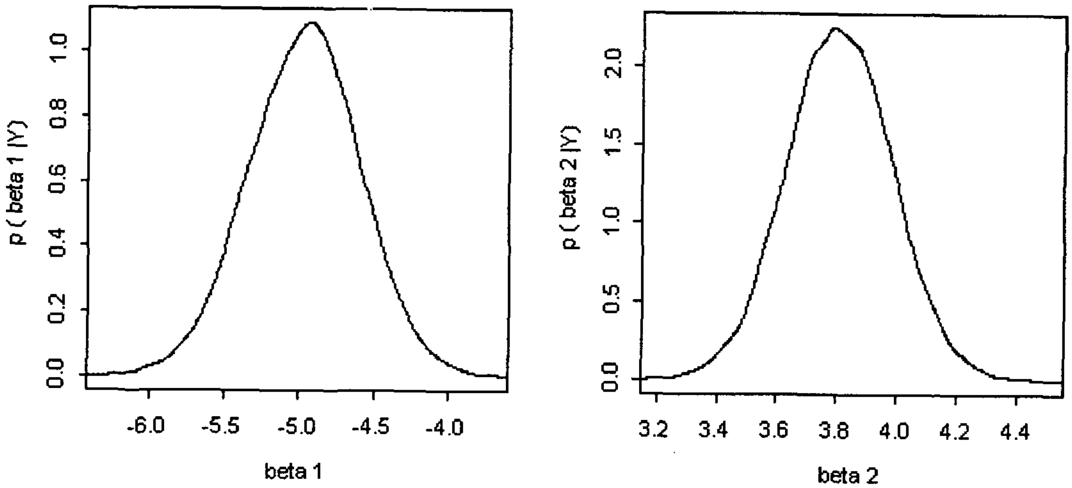

Figure 2 $p(\beta_1|Y)$**(left) and** $p(\beta_2|Y)$**(right) When** n**=32**