Advances and Applications in Statistics
© 2014 Pushpa Publishing House, Allahabad, India
Published Online: January 2015
Available online at http://pphmj.com/journals/adas.htm
Volume 43, Number 1, 2014, Pages 37-52

A BAYESIAN METHOD TO ESTIMATE THE PARAMETERS OF REGRESSION MODEL

Nasr Rashwan and Hanaa Salem

Faculty of Commerce
Tanta University
Egypt
e-mail: nasrrashwan@yahoo.com
hanaa_s2000@yahoo.com

Abstract

This study presents the Bayesian method to estimate the parameters of regression model as an alternative method to the classical methods. Although, this method includes complex calculations but it is more accurate than other methods because it produces a direct probability statement about parameters and allows one to interpret a probability as a measure of degree of belief concerning the actual observed data. In Bayesian method, Gibbs sampling algorithm was used to select several iterations (samples) from posterior distribution of regression coefficients using noninformative prior and conjugate prior. The empirical results showed that the estimates of coefficients under the Bayesian regression model using conjugate prior are more accurate than the Bayesian regression model using noninformative prior.

Introduction

The Bayesian approach to statistical analysis is different from the usual "classical" approach. In the classical approach, the data are only source of

Received: September 7, 2014; Accepted: October 29, 2014

2010 Mathematics Subject Classification: 62-XX.

Keywords and phrases: Bayesian method, noninformative prior, conjugate prior, Gibbs sampling, posterior distribution.

information explicitly taken into account in constructing an estimate or test. In the Bayesian approach, an estimate or test is produced by combining the current data with information from past experience.

The basic formula that is used to incorporate past knowledge into statistical analysis was discovered by Tomas Bayes around 1760. Perhaps the first application to regression problems was made by Harold Jeffreys in 1939.

Classical approach assumes that the parameters are fixed, unknown quantities, and the observed data *y* can be treated as random variable. The goal of the classical approach is to produce estimate of these unknown parameters. These estimates are denoted by $\hat{\theta}$. The most common way to obtain these estimates is by the maximum likelihood method.

When performing Bayesian inference, the assumptions are quite different. The unknown parameters θ are treated as random variables, while the observed data *y* is treated as fixed, known quantities.

The primary goal of each Bayesian analysis is to build a model for the relationship between unknown parameter θ which it may be a single parameter or a vector of many parameters and the data set of *x* which it may be a vector of observations of a single variable or a matrix with multiple observations of many variables.

From Likelihood to Bayesian analysis

The strength of maximum likelihood estimation relies on its large-sample properties, namely that when the sample size is sufficiently large, we can assume both normality of the test statistic about its mean and that likelihood ratio tests follow χ^2 distributions. These nice features do not necessarily hold for small samples.

An alternative way to proceed is to start with some initial knowledge about the distribution of the unknown parameter(s), $p(\theta)$. From Bayes' theorem, the data (likelihood) augment the prior distribution to produce a posterior distribution,

$$P(\theta\backslash x) = \frac{1}{P}\frac{P(x\backslash\theta)P(\theta)}{P(x)} \tag{1}$$

$$= (\text{normalizing constant}) \cdot P(x\backslash\theta)P(\theta) \tag{2}$$

$$= \text{Constant} \cdot \text{likelihood} \cdot \text{prior}. \tag{3}$$

As $P(x\backslash\theta) = L(\theta\backslash x)$ is just the likelihood function. $\frac{1}{P(x)}$ is a constant (with respect to θ), because our concern is the distribution over θ. Because of this, the posterior distribution is often written as

$$P(\theta\backslash x) \propto L(\theta\backslash x)P(\theta). \tag{4}$$

The constant $P(x)$ normalizes $P(x\backslash\theta)$. $P(\theta)$ to one, and hence can be obtained by integration,

$$P(x) = \int_{\theta} P(x\backslash\theta)P(\theta)d\theta. \tag{5}$$

The dependence of the posterior on the prior (which can easily be assessed by trying different priors) provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected under different priors, the data are likely highly informative. To see this, taking log on equation (3) gives

$$\text{Log}(\text{posterior}) = \log(\text{likelihood}) + \log(\text{prior}). \tag{6}$$

A very strong feature of Bayesian analysis is that we can remove the effects of the nuisance parameters by simply integrating them out of the posterior distribution to generate a marginal posterior distribution for the parameters of interest (for more details, see Lee [6], Walsh [10]),

$$P(\theta_1) = \int_{\theta_n} P(\theta_1, \theta_n\backslash y)d\theta_n. \tag{7}$$

The prior distribution is a key part of Bayesian inference and represents the information about an uncertain parameter θ that is combined with the probability distribution of new data to yield the posterior distribution. The use of a prior distribution depends on guess theory and can vary from one statistician to another. There are two types of prior distributions: the first type; called *noninformative prior*, is used when we have very little knowledge or information about the prior distribution (no prior information). The simplest and oldest rule for determining a noninformative prior is the principle of indifference which assigns equal probabilities to all possibilities. Thus, the choice of most frequently used and common of the noninformative prior is a uniform prior (a flat distribution). The uniform prior distribution is improper that is, the function used as a prior probability density has an infinite integral but this problem is not necessarily serious if the likelihood has abounded range such that the function is nonzero.

The second type, called the *conjugate prior*, is one that combines with the distribution if the data vector to yield a posterior distribution that has the same form as the prior distribution (for more details, see Birkes and Dodge [1], Gelman et al. [5] and Stauffer [8]).

Bayesian Simple Linear Regression Using Noninformative Prior

It is considered the following simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_1 + e_i, \quad i = 1, 2, ..., n, \tag{8}$$

where β_0 and β_1 are unknown parameters, $\{y_i\}_{i=1}^n$ are values of the response variable y, $\{x_i\}_{i=1}^n$ are values of the predictor variable x and $\{e_i\}_{i=1}^n$ are normal distribution random errors with zero mean and variance σ^2. The model that relates observations and parameters is written

$$(y, x \backslash \beta_0, \beta_1, \sigma^2) \sim N(\beta_0 + \beta_1 x_i, \sigma).$$

Then the likelihood function becomes

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right). \tag{9}$$

Assuming σ in known, the noninformative prior distribution that commonly used for linear regression is

$$P(\beta_0, \beta_1) \propto 1. \tag{10}$$

The posterior density is

$$P(\beta_0, \beta_1 \backslash y, x, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left(\frac{-\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right). \tag{11}$$

We take the estimates of β_0 and β_1 to be $E(\beta_0)$ and $E(\beta_1)$ under the posterior distribution. These estimates are unbiased, that is $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$.

Bayesian Simple Linear Regression Using Conjugate Prior

Let $y_1, ..., y_n$ be a simple random sample with normal density function as follows:

$$f(y \backslash \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{-(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right). \tag{12}$$

The likelihood of $y_1, ..., y_n$ is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_i - (\beta_0 - \beta_1 x_i))^2}{2\sigma^2} \right). \tag{13}$$

Assume that $P(\beta_0, \beta_1)$ is the joint prior distribution and since β_0 and β_1 are independent, $P(\beta_0, \beta_1)$ is the product of two individual priors (Eltelbany [3]).

As result, the joint posterior distribution is proportional to the product of the likelihood with the joint prior distribution

$$P(\beta_0, \beta_1 \backslash y, \sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right) \cdot P(\beta_0, \beta_1), \quad (14)$$

β_0 is a random variable with normal distribution with mean μ_1 and variance δ_1^2,

$$P(\beta_0) = \frac{1}{\sqrt{2\pi}\delta_1} \exp\left(\frac{-(\beta_0 - \mu_1)^2}{2\delta_1^2} \right). \quad (15)$$

β_1 is a random variable with normal distribution with mean μ_2 and variance δ_2^2,

$$P(\beta_1) = \frac{1}{\sqrt{2\pi}\delta_2} \exp\left(\frac{-(\beta_1 - \mu_2)^2}{2\delta_2^2} \right). \quad (16)$$

The joint prior distribution is

$$P(\beta_0, \beta_1) = P(\beta_0) \cdot P(\beta_1). \quad (17)$$

The posterior distribution

$$P(\beta_0, \beta_1 \backslash y, \sigma^2)$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right) \cdot P(\beta_0, \beta_1)}{\int_\theta \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right) \cdot P(\beta_0, \beta_1) d\theta}, \quad (18)$$

where $\theta = [\beta_0, \beta_1]$.

The marginal posterior distributions for the parameters are:

$$P(\beta_0 \backslash y, \sigma^2) = \int_{\beta_1} P(\beta_0, \beta_1, \sigma^2 \backslash y) d\beta_1, \quad (19)$$

$$P(\beta_1 \backslash y, \sigma^2) = \int_{\beta_0} P(\beta_0, \beta_1 \backslash \sigma^2, y) d\beta_0. \quad (20)$$

Bayesian Multiple Regression Using Noninformative Prior

Now, we consider the regression model in which a response variable y is related to one or more explanatory or predictor variables $x_i, x_2, \ldots, x_{k-1}$ for a random sample of n observations, the model is

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + e_i.$$

Also, this model can be written in matrix form as

$$Y = XB + e, \tag{21}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n*1}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n_1} & x_{n_2} & \cdots & x_{nk-1} \end{bmatrix}_{n \times k},$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n \times 1}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix}_{n \times 1}.$$

Based on this normal model, the likelihood is

$$L = (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right], \tag{22}$$

where T refers to transpose of matrix.

Consider the noninformative prior distribution

$$P(\beta) \propto 1. \tag{23}$$

Then posterior distribution is

$$\begin{aligned} P(\beta \backslash X, \sigma^2) &= \text{Likelihood} * \text{Prior distribution} \\ &= L * 1. \end{aligned} \tag{24}$$

Bayesian Multiple Regression Model Using Conjugate Prior

Let $y_1, \ldots, y_n$ be a simple random sample with normal distribution as follows:

$$f(Y/\beta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - X\beta)^T (y - X\beta)}{2\sigma^2}\right). \tag{25}$$

The likelihood function is:

$$L = \prod_{i=1}^{n} P(y_i \backslash x_i, \beta, \sigma^2)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right), \tag{26}$$

where β is a random vector has a multivariate normal distribution with mean vector μ and variance matrix $\sum$, so the density function of β is

$$P(\beta) = (2\pi)^{-P/2} |\sum|^{-1/2} \exp\left(\frac{-1}{2}(\beta - \mu)^T \sum^{-1}(\beta - \mu)\right). \tag{27}$$

The posterior function is

$$P(\beta \backslash y, x, \sigma^2) = \frac{P(\beta) \cdot L}{P(y)} \tag{28}$$

or

$$P(\beta \backslash y, x, \sigma^2) \propto P(\beta) \times L. \tag{29}$$

This posterior function has exactly the same shape a normal distribution, with

$$E(\beta \backslash y, x, \sigma^2) = (\Sigma^{-1} + x^T x/\sigma^2)^{-1} (\Sigma^{-1} \mu + x^T y/\sigma^2) \tag{30}$$

and

$$V(\beta \backslash y, x, \sigma^2) = (\Sigma^{-1} + x^T x/\sigma^2)^{-1}. \tag{31}$$

Gibbs Sampling

Gibbs sampling algorithm has been widely used on broad class of areas, e.g., Bayesian analysis, statistical inference, bioinformatics and econometrics. It is a particular form of Markov Chain Monte Carlo (MCMC) algorithm for approximating Joint and Marginal distributions by sampling from conditional distributions.

In the case of multiple regression model, the calculations are more complicated and more difficult to evaluate and analyze the solution of a posterior distribution. The Gibbs sampling algorithm becomes quite useful, it is a technique that generates random variables from a distribution without having to calculate the density.

The Gibbs sampling can be used to compute the posterior distribute of β and allows one to sample from a multivariate distribution using full conditional distribution of a parameter given all the other parameters in the model (Gelfand and Smith [4]). Tektas and Guany [9] presented the Gibbs sampling algorithm for the parameter β_j $(j = 0, 1, \ldots, k)$ as follows: The initial values are taken as $(\beta_0^{(0)}, \beta_1^{(0)}, \ldots, \beta_k^{(0)})$.

After the initial values of the parameters are determined, Gibbs sampling is implemented by sampling from the full conditional distributions:

$$\begin{aligned} &\beta_0^{(1)} \text{ from } \pi(\beta_0 \backslash \beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_k^{(0)}, y, x) \\ &\beta_1^{(1)} \text{ from } \pi(\beta_1 \backslash \beta_0^{(1)}, \beta_2^{(0)}, \ldots, \beta_k^{(0)}, y, x) \\ &\vdots \\ &\beta_k^{(1)} \text{ from } \pi(\beta_k \backslash \beta_0^{(1)}, \beta_1^{(1)}, \ldots, \beta_{k-1}^{(1)}, y, x). \end{aligned} \tag{32}$$

Suppose we draw *m* data vectors $\beta^{(L)} = (\beta_0^{(L)}, \beta_1^{(L)}, \ldots, \beta_x^{(L)})$, $L = 1, 2, \ldots, m$ from posterior distribution. According to Gelfand and Smith [4], we can find accurate Bayesian estimate for vector β as

$$\hat{\beta}_{\text{Bayes}} = \frac{\sum_{i=1}^{m} \beta^{(L)}}{m}. \tag{33}$$

Results and Conclusion

In this section, we will focus on the analysis of multiple linear regression using the Bayesian method as an alternative method of least squares estimates. To simplify our complexity calculations WinBugs 1.4 of the Gibbs sampling software was used to compute the regression coefficients as follows (Spiegelhalter et al. [7]):

- In Bayesian multiple linear regression model using noninformative prior, given that we have no prior knowledge of the data in Table 1. For this reason, we will use a uniform prior distribution, $P(\beta_0, \beta_1, \beta_2) \propto 1$ and variance is constant value which is assumed to be 0.001.

Phosphorus Data Example

An investigation of the source from which corn plants obtain their phosphorus was carried out. Concentrations of phosphorus in parts per millions in each of 18 soils were measured. The variables are:

$x_1 =$ concentrations of inorganic phosphorus in the soil.

$x_2 =$ concentrations of organic phosphorus in the soil, and

$y =$ phosphorus content of corn grown in the soil at 20°C.

This example is taken from Chatterjee and Hadi [2, p. 81].

Table 1. Corn plants data

y	64	60	71	61	54	77	81	93	93	51	76	96	77	93	95	54	168	99
x_1	0.4	0.4	3.1	0.6	4.7	1.7	9.4	10.1	11.6	12.6	10.9	23.1	23.1	21.1	23.1	1.9	26.2	29.9
x_2	53	23	19	34	24	65	44	31	29	58	37	46	50	44	56	36	58	51

By using the likelihood function in equation (22) and using the program is showed in the following Figure 1:

1) program code

```
model
{
for(i in 1:n)
{y[i] ~ dnorm(mu[i],0.001)
mu[i]<- beta0 +beta1*x1[i] +beta2*x2[i]
}
beta0 ~ dunif(0,0.001)
beta1 ~ dunif(0,0.001)
beta2 ~ dunif(0,0.001)
}
```

2. data

```
list(x1=c(0.4,0.4,3.1,0.6,4.7,1.7,9.4,10.1,11.6,12.6,10.9,23.1,23.1,21.1,23.1,1.9,26.8,29.9
),
x2=c(53.0,23.0,19.0,34.0,24.0,65.0,44.0,31.0,29.0,58.0,37.0,46.0,50.0,44.0,56.0,36.0,58.
0,51.0),
y=c(64.0,60.0,71.0,61.0,54.0,77.0,81.0,93.0,93.0,51.0,76.0,96.0,77.0,93.0,95.0,54.0,168.
0,99.0),
n=18)
```

3. initial values

```
list(beta0=0.000000001, beta1=0.0000000001, beta2=0.00000001)
```

Figure 1. WinBugs program (noninformative prior).

After performing several iterations (60,000 iterations), we obtain the marginal posteriors densities of parameters β_0, β_1 and β_2 in Figure 2 and a summary of results in Figure 3.

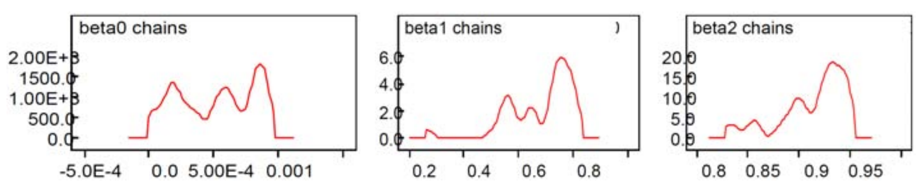

Figure 2. The posterior distribution of estimates.

node	mean	sd	MC error	2.5%	median	97.5%
beta0	5.222E-4	2.996E-4	8.634E-5	9.027E-6	5.842E-4	9.401E-4
beta1	0.6854	0.1194	0.03199	0.2681	0.7366	0.8222
beta2	0.9143	0.03289	0.008113	0.8301	0.926	0.9516

Figure 3. Summary results in WinBugs.

From the previous results, the posterior means are serve as estimates of regression

$$E(\beta_0 \backslash Y,\ X) = 0.00052, \quad E(\beta_1 \backslash Y,\ X) = 0.6854,$$

$$E(\beta_1 \backslash Y,\ X) = 0.9143$$

and the regression model is

$$\hat{y} = 0.00052 + 0.6854x_1 + 0.9143x_2.$$

Note that, in the case of noninformative prior, the posterior means are approximately equal to the maximum likelihood estimates of the parameters $(\beta_0, \beta_1 \text{ and } \beta_2)$.

- In Bayesian multiple linear regression using conjugate prior we assumed that the parameters (β vector) have multivariate normal distribution with mean vector (μ) and variance-covariance matrix $\sum$ as showed in equation (27). By using the likelihood equal (26) and we use the program in Figure 4.

1) program code

```
model
{
for(i in 1:n)
{y[i] ~ dnorm(mu[i],0.01)
mu[i]<- beta0 +beta1*x1[i] +beta2*x2[i]
}
beta0 ~ dnorm(0,0.000000000001)
beta1 ~ dnorm(0,0.000000000001)
```

beta2 ~ dnorm(0,0.000000000001)

}

2. data

list(x1=c(0.4,0.4,3.1,0.6,4.7,1.7,9.4,10.1,11.6,12.6,10.9,23.1,23.1,21.1,23.1,1.9,26.8,29.9),

x2=c(53.0,23.0,19.0,34.0,24.0,65.0,44.0,31.0,29.0,58.0,37.0,46.0,50.0,44.0,56.0,36.0,58.0,51.0),

y=c(64.0,60.0,71.0,61.0,54.0,77.0,81.0,93.0,93.0,51.0,76.0,96.0,77.0,93.0,95.0,54.0,168.0,99.0), n=18)

3. initial values

list(beta0=0, beta1=0, beta2=0)

Figure 4. WinBugs program.

After performing iterations, we obtain the marginal posterior distributions of parameters in Figure 5 and a summary of some the Bayesian results in Figure 6.

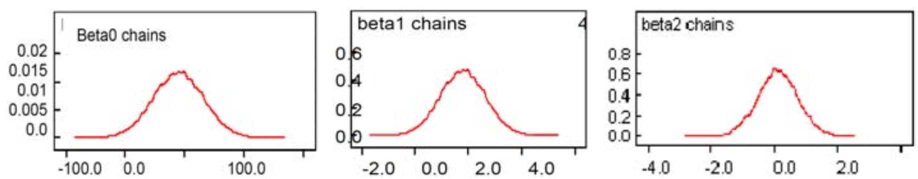

Figure 5. The posterior distributions of estimates.

node	mean	sd	MC error	2.5%	median	97.5%
beta0	55.63	6.647	0.006376	43.23	56.34	69.94
beta1	1.839	0.02647	0.0003982	1.354	1.848	2.62
beta2	0.09477	0.01411	0.0002161	-0.1214	0.1027	0.4221

Figure 6. Summary results in WinBugs.

From the previous, we find that the estimate of vector of regression parameters which referred to a posterior means and are given by

$$E(\beta_0 \backslash Y,\ X,\ \sigma^2) = 55.63, \quad E(\beta_1 \backslash Y,\ X,\ \sigma^2) = 1.839,$$

$$E(\beta_2 \backslash Y,\ X,\ \sigma^2) = 0.0947.$$

This means that $\hat{\beta} = \begin{bmatrix} 55.63 \\ 1.839 \\ 0.0947 \end{bmatrix}$ and the regression model is given by

$$\hat{y} = 55.63 + 1.839x_1 + 0.0947x_2.$$

Testing the significance of intercept and regression coefficients, we can test $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ at significance level α by using the appropriate critical value.

If the null hypothesis is true, then the population of all possible values of the test statistic

$$t = \frac{b_j - \beta_j}{sd_j}$$

has a t distribution with $n - (k + 1)$ degrees of freedom.

Table 2

Noninformative				**Conjugate**		
j	b_j	sd_j	t	b_j	sd_j	t
0	5.222E-4	2.996E-4	1.74	55.63	6.647	8.37
1	0.6854	0.1194	5.74	1.839	0.02647	69.47
2	0.9143	0.03289	27.79	0.09477	0.01411	6.716

Table 2 summarizes the calculation of the t statistics for testing the significance of the intercept and regression coefficient. Since in noninformative case, we reject $H_0 : \beta_j = 0$ for $j = 1, 2$ at the 0.05 level of significance, we use the critical value $t_{0.025,15} = 2.131$ and the intercept is not significant. In conjugate case, we reject $H_0 : \beta_j = 0$ for $j = 0, 1, 2$ at the 0.05 level of significance.

- Comparison between the Bayesian multiple regression model using noninformative prior and the Bayesian multiple regression model using conjugate prior, in the following Table 3 which shows some of the Bayesian results from 60,000 iterations obtained from posterior distributions of the regression parameters.

Table 3. Some Bayesian results

Coefficients	Bayesian regression (noninformative prior)			Bayesian regression (conjugate prior)		
	Mean	sd_j	MC error	Mean	sd_j	MC error
β_0	0.00052	0.00029	0.00086	55.63	6.647	0.00637
β_1	0.6854	0.1194	0.03199	1.839	0.02647	0.000398
β_2	0.9143	0.0329	0.00811	0.0947	0.0114	0.000212

From Table 3, we find that

- The standard deviation (sd) of regression coefficients (β_1, β_2) in the case of conjugate prior is less than the standard deviation of regression coefficients (β_1, β_2) in the case of noninformative prior.

- Also, the Markov Chain error (MC error) in the conjugate prior is less than the MC error in the noninformative prior.

Based on the previous results, we infer that the Bayesian regression model under conjugate prior is better and more accurate than the Bayesian regression model under flat noninformative prior.

References

[1] D. Birkes and C. Dodge, Alternative Methods of Regression, John Wiley and Sons, Inc., New York, 1993.

[2] S. Chatterjee and A. S. Hadi, Sensitivity Analysis in Linear Regression, John Wiley and Sons, Inc., New York, 1988.

[3] D. Eltelbany, Bayesian regression with missing data, M.Sc. Thesis, Faculty of Commerce, Benha University, Egypt, 2012.

[4] A. E. Gelfand and A. F. Smith, Sampling-based approaches to calculating marginal densities, J. Amer. Statist. Assoc. 85 (1990), 398-409.

[5] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, Bayesian Data Analysis, 2nd ed., Chapman and Hall, London, 2003.

[6] P. M. Lee, Bayesian Statistics: An Introduction, Oxford University Press, New York, 1989.

[7] D. J. Spiegelhalter, A. Thomas and N. G. Best, WinBugs version 1.4 user manual, MRC Biostatistics unit, 1999. Software Available at: http://www.mrc-bsu.com.ac.uk/bugs/winbugs/contents.shtml.

[8] H. B. Stauffer, Contemporary Bayesian and Frequentist Statistical Research Methods for Natural Resource Scientists, John Wiley and Sons, Inc., New York, 2008.

[9] D. Tektas and S. Gunay, A Bayesian approach to parameter estimation in binary logit and probit models, J. Math. Stat. 37 (2008), 167-176.

[10] B. Walsh, Introduction to Bayesian analysis, Lecture Notes for EEB596, 2002. Available at: http://nitro.biosci.arizona.edu/courses/EEB596/handouts/Bayesian.