ISSN 1684-8403
Journal of Statistics
Volume 21, 2014. pp. 62-74

Bayesian Inference for Logit-Model using Informative and Non-informative Priors

Tahir Abbas Malik[1] and Muhammad Aslam[2]

Abstract

In the field of econometrics analysis of binary data is widely done. When the sample is small Bayesian approach provides more appropriate results on classical approach (MLE). In Bayesian approach the results can be improved by using different Priors. It is known that the binary data can be modeled by using Logistic, Probit or Tobit Links. In our study, we use Logistic Link. When modeling binary data, the shape of the distribution of Regression coefficients is no more normal. To find the coefficient of Skewness of Regression coefficients we use different Priors. It is observed that Haldane Prior provides better results than Jeffreys Prior, while Informative Prior performs better than the other Non-informative Priors for the data set under consideration.

Keywords

Polychotomous Response, Binary Logit model, Non-informative Prior, Bayesian analysis, Skewness

1. Introduction

If the dependent variable in a data set is categorical, it is not possible to use Linear Regression Models to estimate the unknown Regression coefficients. So, a Generalized Linear Model (GLM) is an alternative technique to estimate the unknown parameters.

[1] Department of Statistics, Government College University, Lahore, Pakistan.
Email: malik83qau@yahoo.com
[2] Department of Statistics, Quaid-i-Azam University, Islamabad 44000, Pakistan.
Email: aslamsdqau@yahoo.com

Since we know that while using Maximum Likelihood method for the estimation of Regression coefficients it may mislead when we have small sample data sets, because MLE are usually based on Asymptotic Theory. Griffiths et al. (1987) found that MLE have significant Bias for small samples. This problem can be handled by using Bayesian Technique while estimating Regression parameters.

Let us consider here that the response variable y_i is categorical in nature with binary options coded as [0 or 1]. It is obvious that y_i follows a Bernoulli Distribution where $y_i=1$ with probability p_i and $y_i=0$ with probability $1-p_i$. Thus $E(y_i) = p_i$ and $Var(y_i) = p_i(1-p_i)$. Let $y = (y_1, y_2, ..., y_n)'$ be a sample of n observations. Then for a sample of n observations the Likelihood Function is,

$$L\left(data \mid parameter\right) = \prod_{i=1}^{n} \left\{ p_i^{\,y_i} (1-p_i)^{1-y_i} \right\} \tag{1.1}$$

Now if $p_i = H\left(\beta x_i'\right)$ then the Joint Likelihood Function could be written as:

$$L\left(data \mid \beta\right) = \prod_{i=1}^{n} \left\{ H(x_i'\beta)^{y_i} (1-H(x_i'\beta))^{1-y_i} \right\} \tag{1.2}$$

Here β is the vector of Regression coefficients and x_i' the set of explanatory variables. While H is the Link Function that is Logistic in our case and we will use this Link throughout our study to obtain the Posterior estimates. The unknown β's are considered independent here, so the Point Posterior Distribution can also be written as:

$$p\left(\beta \mid data\right) \propto \prod_{i=1}^{n} \left\{ H(x_i'\beta)^{y_i} (1-H(x_i'\beta))^{1-y_i} \right\} \times p\left(\beta\right) \tag{1.3}$$

The data set for our study is taken from Cengiz et al. (2001). The data set contains the sample observations of 32 individuals. This research was actually made by the Institute of Medical Research, Kuala Lumpur, Malaysia. They used Erythrocyte Sedimentation Rate (ESR) related to two plasma proteins, fibrinogen and Y-globulin, both measured in gm/l, for a sample of thirty-two individuals. We have classified our variables as follows:

$y_i =$ The Erythrocyte Sedimentation Rate (ESR)

$x_{fi} =$ The amount of protein plasma fibrinogen

$x_{gi} =$ The amount of protein plasma Y-globulin

Now, we will modify the above general form with respect to the variables and we will consider the Binary Logistic Regression Model with two explanatory variables as:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi} \tag{1.4}$$

Here β_0 is the intercept while β_1 and β_2 are the slope coefficients for the independent explanatory variables fibrinogen and Y-globulin, respectively. The above Logistic Regression Model can also be represented as:

$$p_i = p_r(y_i = 1) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}\right)\right)} \tag{1.5}$$

Then the Joint Posterior Distribution of the parameters β_0, β_1 and β_2 are defined as:

$$p\left(\beta_0, \beta_1, \beta_2 \mid data\right) \propto L\left(data \mid \beta_0, \beta_1, \beta_2\right) \times p\left(\beta_0, \beta_1, \beta_2\right) \tag{1.6}$$

Let $\beta = \left(\beta_0, \beta_1, \beta_2\right)$ then the above Joint Posterior Distribution can be modified as:

$$p\left(\beta \mid data\right) \propto L\left(data \mid \beta\right) \times p\left(\beta\right) \tag{1.7}$$

The ESR is a non-specific marker of illness. ESR is the rate at which the red blood cells settle out of suspension in a blood plasma, when measured under standard conditions. The original data were presented by Collett and Jemain (1985) and were reproduced by Collett (1996), who classified the ESR as binary (0 or 1). Since the ESR for a healthy individual should be less than 20 mm/h and the absolute value of ESR is relatively unimportant, a response of zero signifies a healthy individual (ESR $<$ 20) while a response of unity refers to an unhealthy individual (ESR $\geq$ 20).

The aim of study is to determine the unknown Regression parameters by Bayesian approach using Informative and Non-informative Priors; all the unknown parameters are assumed to be independent in our study. As many writers used these data i.e. Cengiz et al. (2001), Collett (1996) and Collett and Jemain (1985), but say nothing about the shape of the Distribution of the Regression parameters. We consider this case here and analyze possible change in the coefficient of Skewness by using different Priors. We have also determined the probability of rejecting the null hypothesis which is very simple case while we are stumble upon with Bayesian approach. In proceeding sections we obtain the Posterior modes, Posterior means, standard error and Odds Ratio to check impact on the Regression estimates.

Logistic Regression is universally used when the independent variables include both numerical and nominal measures and the outcome variable is binary, having two values. It requires no assumptions about the Distribution of the independent variables. Another advantage is that, the Regression coefficient can be interpreted in term of Odds Ratio. Bayesian Analysis for Logistic Regression Models is broadly used in modern days. Bayesian Analysis conditions on the data and integrates over the parameters to evaluate the probabilities rather than estimates.

Here, we present different Priors which are suitable for given data set. The standard choice over the recent years has been the invariant Prior anticipated by Jeffreys (1961) that is Jeffrey's Prior. We have also presented Uniform and Improper (Haldane) Prior which are described by Bernardo and Smith (1994). For Informative Prior we have assumed Normal Prior by suggesting suitable values of hyper-parameters. The evaluations of integrals remain very difficult in these situations and hard to study it analytically but numerical methods can be used to overcome this difficulty.

2. Selection of Priors

A Prior may be declared as an Achilles heel (Weakness) of Bayesian Statistics, as in Bayesian Statistics parameters are assumed random. The Priors carry certain Prior information about the unknown parameter(s) that is coherently incorporated into the Inference via the Bayes Theorem. Choice of the Prior Distribution depends upon the nature and the range of the parameter(s) being studied through the Bayesian Analysis. Sometimes, it happens that the Prior elicitation becomes difficult, or a little Prior information is available, then it is conventional to choose Priors which may reflect little Prior information. Such Priors are termed as the Non-informative Priors, indifferent, ignorant, and vague or reference Priors. e.g. Bernardo (1979), Cengiz (2001), Ghosh and Mukerjee (1992), Jeffreys (1961), Kass and Wasserman (1996) and Tibshirani (1989). Here in our study we have suggested three Non-informative Priors and an Informative Prior.

Non-informative Prior, by which we mean a Prior that usually, contains no Prior information about a parameter. The simplest situation is to assign each element Uniform probability, the Uniform Prior.

$$p(\beta) \propto 1 \qquad -\infty < \beta < \infty \tag{2.1}$$

Another very important Improper Prior is Haldane Prior. Haldane Prior is the Improper Dirichlet Distribution having all hyper-parameters to be zero. It was first suggested by Rubin (1981). So, it can be said that if Beta is used as a Prior Distribution with both the parameters equal to zero then the Beta Prior will be Haldane Prior. It can also be derived from Bernoulli Distribution if the Response variables have only two categories as yes or no. So, the Haldane Prior will be as:

$$p_H(\beta) \propto p_i^{-1}(1-p_i)^{-1} \tag{2.2}$$

This Prior can also be derived by partially differentiating the Log-likelihood Function of Sampling Distribution of dependent variable. As we have already defined in the Logistic Model the value of p_i as:

$$p_i = \frac{1}{1+\exp(-\beta_0 - \beta_1 x_{fi} - \beta_2 x_{gi})} \tag{2.3}$$

Jeffrey's (1946, 1961) proposes a Non-informative Prior Distribution. This Prior can also be obtained when we the values of parameter as (1/2, 1/2) in Beta Distribution. These Priors are extensively used in literature as far as the study of Logistic Analysis is concern as they are the Conjugate Prior for Sampling Distribution of Response variable. The Jeffrey's Prior can be defined as:

$$p_J(\beta) \propto p_i^{-1/2}(1-p_i)^{-1/2} \tag{2.4}$$

In Bayesian econometrics, the Normal Prior for unknown parameter of the Regression Model is extensively used, but while studying Logistic Regression Models the use of Normal Prior is not as simple as in Simple Regression Models due to the complicated form of Posterior Distribution as it is not a Conjugate Prior for this Model, usually it is used for scale parameter but in our case we have consider the impact of both the parameters on the Posterior parameters by considering independent Priors for each parameter i.e.

$$p\left(\beta_j;\mu_j,\delta_j\right) = \prod_{j=0}^{2} \frac{1}{\sqrt{2\pi}\delta_j} \exp\left\{-\frac{1}{2}\left(\frac{\beta_j-\mu_j}{\delta_j}\right)^2\right\} \tag{2.5}$$

3. Posterior Analysis

Many applied econometric analyses are concerned with Modeling discrete variables such as yes or no, in or out of labor force, married or unmarried etc. The basic Model in this area is the Dichotomous Response Model, very similar to Bioassay Dosage Response Models. We have the response variable y_i.

3.1 Posterior Distribution using Uniform Prior: The Uniform Prior is very useful Prior when there is lake of Prior information about the parameters to be estimated, in literature the use of Uniform Prior is very wide i.e. Bernardo (1979), Ghosh and Mukerjee (1992), Jeffreys (1961), Kass and Wasserman (1996) and Tibshirani (1989), propose the Bayesian Analysis of unknown parameters using one of the most widely used Non-informative Priors, that is, a Uniform (possibly Improper) Prior that routinely used by Laplace (1812). The Joint Posterior Distribution using Joint Uniform Prior is given as:

$$p\left(\beta \mid data\right) = \frac{1}{k}\prod_{i=1}^{n}\left\{\left(\frac{\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{y_i}\left(\frac{1}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{1-y_i}\right\} \quad (3.1.1)$$

3.2 Posterior Distribution using Jeffrey's Prior: Jeffrey's (1961) proposes a Non-informative Prior. Berger (1985) argues that Bayesian Analysis using Non-informative Prior is the single most powerful method of statistical analysis. The main feature of Jeffrey's Prior is that it is a Uniform measure in Information Metric, which can be regarded as the natural Metric for statistical inference. When there is no information available for the Regression parameters then usually Jeffrey's Prior works as the appropriate choice for the Bayesian Analysis. We have considered the Joint Jeffrey's Prior for the parameters to obtain the Joint Posterior Distribution as:

$$p\left(\beta \mid data\right) = \frac{1}{k}\prod_{i=1}^{n}\left\{\left(\frac{\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{y_i-1/2}\left(\frac{1}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{1/2-y_i}\right\} \quad (3.2.1)$$

3.3 Posterior Distribution using Haldane Prior: Another suitable Prior suggested in literature for GLM is Haldane Prior known Improper Prior because for GLM the Beta Prior works a Conjugate Prior for Bernoulli trial and it sometime obtain when Beta has both the parameters equal to zero. Bernardo and Smith (1994) use this Improper Prior for GLM. Here, the Joint Haldane Prior is considered for the Joint Posterior Distribution of the Regression parameters. The Joint Posterior Distribution is given as:

$$p(\beta \mid data)=\frac{1}{k}\prod_{i=1}^{n}\left\{\left(\frac{\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{y_i-1}\left(\frac{1}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{-y_i}\right\} \quad (3.3.1)$$

3.4 Posterior Distribution using Informative Prior: Normal Priors for GLM have been discussed in the literature write by Dellaportas and Smith (1993). Aitchison and Dunsmore (1975) use Normal Priors for multi-parameter cases. Normal Distribution has strong grounds to be used as Prior in Bayesian Econometric Models. We have considered the independent Normal Priors for all the unknown parameters of the Binary Regression Model, which has three unknown parameters. So, the Joint Posterior Distribution for all the parameters is given below:

$$p(\beta \mid data)=\frac{1}{k}\prod_{i=1}^{n}\left\{\left(\frac{\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{y_i}\left(\frac{1}{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})}\right)^{1-y_i}\right\}\times$$

$$\prod_{j=0}^{2}\frac{1}{\sqrt{2\pi}\delta_j}\exp\left\{-\frac{1}{2}\left(\frac{\beta_j-\mu_j}{\delta_j}\right)^2\right\} \quad (3.4.1)$$

We have Joint Posterior Distributions for all the Joint Priors, to proceed further for Joint Informative Prior which is Normal Prior, the selection of appropriate values for the hyper-parameters is very important, while using Bayesian Technique the selection of hyper-parameters usually done with elicitation using Predictive Distributions but some time the Predictive Distributions are very complicated and to obtain hyper-parameters using these Predictive Distribution are very difficult so, the next appropriate technique is to select the values of hyper-parameters using the expert opinion keeping in view the possible variation in the value of Posterior estimates with these hyper-parameters. We have suggested the values of hyper-parameters which are appropriate for this data analysis. Since we know that the Prior Distributions of parameters β_0, β_1 and β_2 are as follows, $\beta_0 \sim N(a_0,b_0)$, $\beta_1 \sim N(a_1,b_1)$ and $\beta_2 \sim N(a_2,b_2)$. A range of values of hyper-parameters suggested and the Posterior estimates are obtained, then the appropriate values of hyper-parameters that are used for further Bayesian Analysis are suggested as, $\beta_0 \sim N(18.95, 5.15)$, $\beta_1 \sim N(4.75, 3.05)$ and $\beta_2 \sim N(3.25, 0.75)$. We have calculated a range of values of hyper-parameters and selected that values which has minimum standard error.

4. Numerical Results

In this section, the numerical results are obtained for all the Posterior Distributions that are obtained in the previous sections; we have used Log-likelihood Function to obtain the Posterior modes, while Quadrature method is used to obtain Posterior mean and standard error.

The results are given in Table 1, this Table provides us the Posterior modes, Posterior means, standard errors, Odds Ratio and Karl-Pearson coefficient of Skewness. It can be observed that the Posterior mode for β_0 is greater than the Posterior mean of β_0 which indicates that the Distribution of this Posterior estimate is negatively Skewed how much it is Skewed, it can be observed from the value of coefficient of Skewness. We have also observed that the Posterior mean of β_1 is greater than the Posterior mode of β_1 which shows that the Distribution of Posterior estimate is positively Skewed, also same pattern is observed for β_2 that also shows a positively Skewed pattern but coefficient of Skewness is much higher than β_1. The values of Odds Ratio are computed to measure the strength of association between ESR and the available quantities of proteins. As we know that Odds Ratio varies from zero to infinity. So if Odds Ratio is equal to one it means there is no association between the two variables and if it is greater than one it means the two variables are associated and its strength of association is measured with the possible value of Odds Ratio. If the value of Odds Ratio is less than one than they are less associated with each other. Here, in given Table, the values of Odds Ratio are greater than one and for β_1 they are higher as compared to β_2, so it can be said that the strength of association between ESR and fibrinogen is much higher than the association of ESR and Y-globulin.

It is also observed that the Karl-Pearson coefficient of Skewness approaches to zero for both parameters with different Priors, i.e. it is higher for Uniform Prior and least for Normal Prior which indicates that the shape of the parameters approaches to Symmetry for Normal Prior. This type of pattern can also be observed in graphical representation of the parameters with different Priors, these graphs are given in Figures 1-12.

The graphs show the shape of the marginal Distributions of the Regression coefficients. In section (a) the graphs are for the Uniform Prior and it can be seen that their shape is not symmetrical while the graph of β_0 see a(1) show a pattern for negatively Skewed Distribution while the slope coefficients β_1 see a(2) and β_2 see a(3) show a pattern of positively Skewed Distribution. This pattern can also be observed for section (b) where we have presented the graphs for intercept and slope coefficients using Jeffreys Prior, the same can be seen in section (c) for Haldane Prior and in section (d) we have the graphs using Informative (Normal) Prior. The shape of Posterior Distributions of the intercepts and slope coefficients are similar but their level of Symmetry varies with different Priors.

4.1 Hypothesis Testing: Hypothesis testing in Bayesian is very simple; here we only find the Posterior probability of accepting the null hypothesis and comment upon the possible significance of the coefficient by integrating the Joint Posterior Distribution upon the parameters i.e. we test the hypothesis here.

$H_0 : \beta_1 \le 0$ Versus $H_1 : \beta_1 > 0$ and $H_0' : \beta_2 \le 0$ Versus $H_1' : \beta_2 > 0$

The Posterior probability for H_0 while testing β_1 is:

$$p_0 = \int_{-\infty}^{0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p\left(\beta_0, \beta_1, \beta_2 \mid data\right) d\beta_0 d\beta_2 d\beta_1$$

While the Posterior probability for H_0' while testing β_2 is:

$$p_0 = \int_{-\infty}^{0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p\left(\beta_0, \beta_1, \beta_2 \mid data\right) d\beta_0 d\beta_1 d\beta_2$$

If we use the Posterior Distribution while using the Normal Prior the Posterior probability of accepting the null hypothesis of $H_0 : \beta_1 \le 0$ is very small that is 0.001368, so we will conclude in favor of alternative hypothesis and say that it has a very high Odd Ratio to play a significant role in effecting ESR. While the probability of accepting the null hypothesis of $H_0' : \beta_2 \le 0$ is little high as compare to above that is 0.030028. So, the Posterior probability indicate that under Bayesian hypothesis criterion there is 3% chance to accept H_0' and we conclude that Y-globulin has a positive effect on ESR but cannot be considered as significant for ESR as the fibrinogen is. The Non-informative Prior also provides the similar conclusion with little deviation in Posterior probabilities.

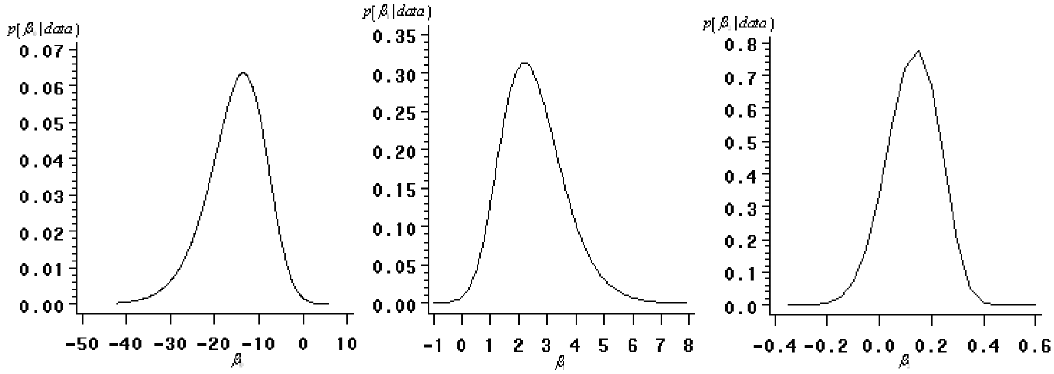
5. Conclusion

The aim of this article is to use Bayesian approach to estimate the Regression parameters of Logit Model and the coefficient of Skewness by using different Priors. The results are more appropriate by using Informative Priors but in case of Non-informative Priors it is observed that Haldane Prior provides better results than Jeffreys Prior for this particular data set. The Odds Ratio provides the evidence for all Priors that fibrinogen effect more on ESR as compare to globulin as its effect is highly significant.

It can also be seen from graphs that while using Normal Prior the shape of the graphs are close to Symmetry as compare to others Non-informative Priors. So it can be concluded that the results obtained through Normal Prior are more appropriate than the results of Non-informative Priors.

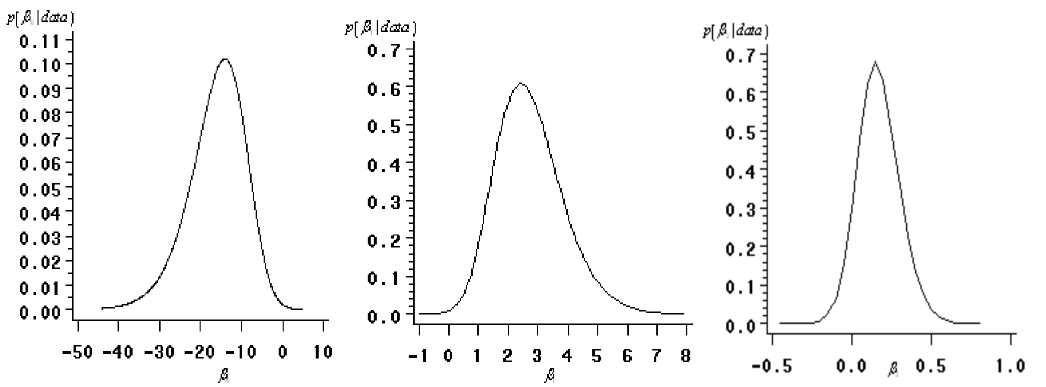
Table 1: Results using Different Priors

Coefficient		**Non informative Prior**			**Informative Prior**
		Uniform Prior	**Jeffrey's Prior**	**Haldane Prior**	
$\hat{\beta}_0$	Posterior Mode	−12.5060	−12.2886	−10.5408	−9.9371
	Posterior Mean	−16.3481	−15.2451	−13.1740	−11.8312
	Standard Error	5.7129	5.1827	4.9012	4.7210
	SK_P	−0.6725	−0.5705	−0.5373	−0.4012
$\hat{\beta}_1$	Posterior Mode	2.0363	1.9506	1.8505	1.6378
	Posterior Mean	3.3451	2.5811	2.1892	1.8841
	Odds Ratio	7.6622	7.0329	6.3630	5.1438
	Standard Error	2.9811	1.9050	1.1935	0.9831
	SK_P	0.4390	0.3310	0.2838	0.2505
$\hat{\beta}_2$	Posterior Mode	0.1452	0.1428	0.1371	0.1336
	Posterior Mean	0.3945	0.2993	0.2015	0.1719
	Odds Ratio	1.1563	1.1535	1.1469	1.1429
	Standard Error	0.3132	0.2187	0.1534	0.1345
	SK_P	0.7960	0.7156	0.4198	0.2848



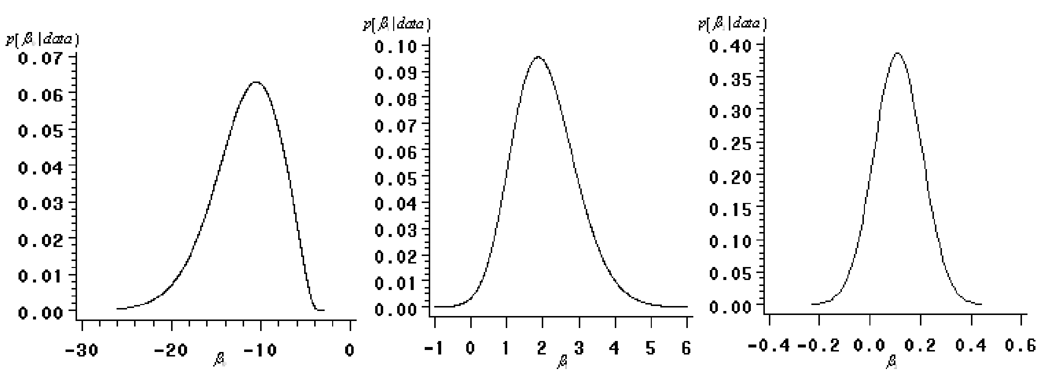
a(1) a(2) a(3)

Figure 1-3: Graphs Posterior Distribution using Uniform Prior



b(1) b(2) b(3)

Figure 4-6: Graphs Posterior Distribution using Jeffrey’s Prior



c(1) c(2) c(3)

Figure 7-9: Graphs Posterior Distributions using Haldane Prior

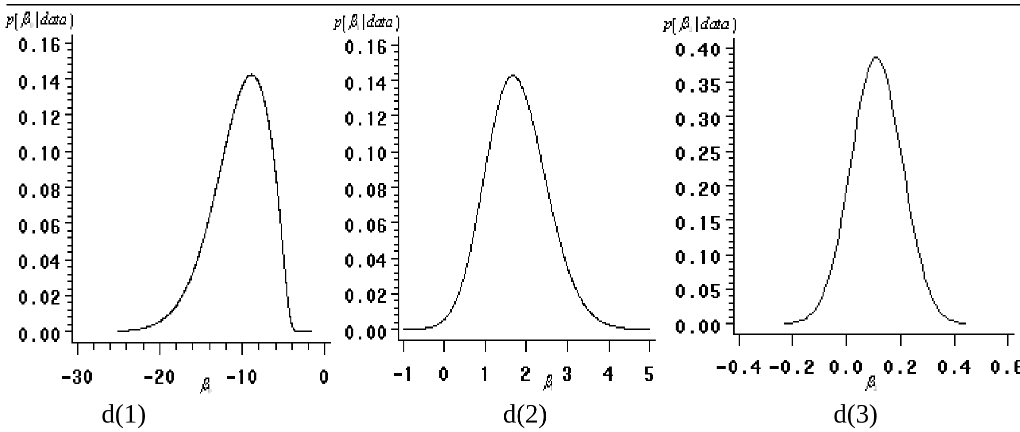


Figure 10-12: Graphs Posterior Distributions using Normal Prior

References

1. Aitchison, J. and I. R. Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge University Press, U.K.
2. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, (Second Edition)*. Springer Verlag, New York.
3. Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society (B),* **41**, 113-147.
4. Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley and Sons, England.
5. Cengiz A. M., Bek Y. and Yilmaz R. (2001). Bayesian Inference of Binary Logistic Regression Model for Assessing Erythrocyte Sedimentation Rate. *Pakistan Journal of Biological Sciences*, **4(9)**, 1180-1183.
6. Collett, D. (1996). *Modeling Binary Data*. Chapman and Hall, London.
7. Collett, D. and Jemain, A. A. (1985). Residuals, outliers and influential observations in Regression Analysis. *Sains Malaysiana*, **14**, 493-511.
8. Dellaportas, P. and Smith, A. F. M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. Journal of Applied Statistics, **42**, 443-459.
9. Ghosh, J. K. and Mukerjee, R. (1992). *Bayesian Statistics (Forth edition).* Oxford University Press, Enagland.

10. Griffiths, W. E., Hill, R. C. and Pope, P. J. (1987). Small sample properties of Probit Model estimators. *Journal of American Statistical Association*, **82**, 929-937.
11. Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, England.
12. Kass, R. E. and Wasserman, L. (1996). The selection of Prior Distributions by formal rules. *Journal of the American Statistical Association*, **91**, 1343-1370.
13. Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, **9(1)**, 130-134.
14. Tibshirani, R. (1989). Non-informative Priors for one parameter of many. *Biometrika*, **76(3)**, 604-608.