Bayesian Logistic Regression Analysis

QUAID-I-AZAM UNIVERSITY

By

Tahir Abbas Malik

Department of Statistics,
Quaid-i-Azam University, Islamabad
2010

In the name of ALLAH,
The Most Beneficent,
The Most Merciful.

Bayesian Logistic Regression Analysis

By

Tahir Abbas Malik

A Thesis Submitted In The Partial Fulfillment Of The Requirement For The Degree Of Master of Philosophy
In

Statistics

Department of Statistics,
Quaid-i-Azam University, Islamabad

2010

Certificate

Bayesian Logistic Regression Analysis

By

Tahir Abbas Malik

A Thesis Submitted In The Partial Fulfillment Of The Requirement For The Degree Of Master of Philosophy
In
Statistics

We accept this thesis as conforming to the required standards.

1. ____________ 3/8/2010
Dr. Javid Shabbir
(Chairman)

2. ____________
Prof. Dr. Muhammad Aslam
(Supervisor)

3. ____________
(External Examiner)

Date. 26/7/2010

Department of Statistics,
Quaid-i-Azam University, Islamabad
2010

DECLARATION

I hereby solemnly declare that the thesis entitled "Bayesian Logistic Regression Analysis" submitted by me for the partial fulfillment of the Master of Philosophy (M.Phil) in statistics, is the original one and has not been submitted concurrently or latterly to this or any other university for any other degree.

Date: 28-07-2010

Signature: Malik

(Tahir Abbas Malik)

Dedicated to

My parents and teachers,

Particularly Professor Dr. Muhammad Aslam,

Who gave me a perfect and affectionate supervision to complete this thesis

ACKNOWLEDGEMENT

First of all I would like to thank to Almighty Allah, The compassionate and the most merciful, who bestowed to me, His blessings to complete this task. All respect to His Prophet (Peace be upon him) for enlightening the essence of faith in Allah, converging all his kindness and mercies upon me.

I am privileged to thank Prof. Dr. Muhammad Aslam for taking up the responsibility as my research supervisor. I offer tribute to him for patience, affection attitude and overall patronization.

I wish to express my warm and sincere thanks to Dr. Javid Shabbir, Chairman Department of Statistics Quaid-i-Azam University Islamabad and the faculty members, Dr. Farid Khan, Dr. Zahid Asghar, Mr. Jamshed Iqbal, Dr. Zawar Hussain, Dr. Muhammad Riaz, Mrs. Maryam Asim, Miss Ameena Urooj for enlightening my statistical know how during course work and guidance during my research work.

Special thanks are extended to my senior fellows Dr. Ahmar Mahmood, Dr. Nasir Abbas and Dr. Muhammad Saleem and my jigars (good friends) Cheema, Hummy, Zafari, Rafique Bhai, Jaji, Shahzad, Manzoor, Younas, Yousif Ali Khan, Aabi, Zeemi, Mazhari, Beloo, Faisee, Sabir, Mohsin, Sajid my all class fellows, and my sweet juniors for their cooperation in all aspects.

I am also thankful to all staff members especially (Khurram, Ikaram Bhai, Moeed Bhai and Nain) of the department who gave me their nice cooperation.

I express my warm and sincere thanks to all the faculty members in CIIT Abbottabad, particularly Dr. Madad Khan HoD Department of Mathematics CIIT Abbottabad, Ossam Bhai and Imran Zulfiqur Cheema for their kind attitude and cooperation during my research work.

Last but not the least, I express my gratitude for my parents, sisters, brothers, cousins (Ghazan Bhai & Malik Muhammad Arif) and particularly to my Mother and elder brother Kausar Abbas Malik whose encouragement, sincerity for my study, affection, prayers and wishes have been a great source of comfort during my whole education period.

(Tahir Abbas Malik)

ABSTRACT

This thesis addresses the Bayesian Analysis of three binary logistic regression models i.e. binary logistic regression model without intercept, with intercept and with two explanatory variables. One informative (Normal) and three noninformative priors are assumed for the parameters of three models. All the analysis is carried out in SAS package. We have selected the hyperparameters for informative prior on basis of expert opinion and use for further analysis. We have used the data set of Erythrocyte Sedimentation Rate (ESR) form Cengiz et al. (2001), that is binary in nature and coded [0, 1] with two explanatory variables: Fibrinogen and Y-globulin that are blood plasma proteins. We have used the logistic link for logistic regression analysis.

We proceed with Bayesian analysis for all the logistic regression models, the noninformative priors are derived, then based on posterior distribution we have obtained the posterior modes, posterior means, posterior standard deviation and Karl Pearson Coefficient of Skewness to say about the shape of the distribution of parameters and the results are further compared with classical results. To check the significance of parameters we have developed programs in SAS package to find posterior probabilities, the Bayesian approach to hypotheses testing has been carried out. The comparison of Bayesian and Classical results is also presented. We have also suggested the appropriate model. Proposed SAS package plays major role for completion of this dissertation.

TABLE OF CONTENTS

Chapter 1

1.1 Introduction to Statistics

Statistics is the science of making effective use of numerical data relating to groups of individuals or experiments. It deals with all aspects of this, including not only the collection, analysis and interpretation of such data, but also the planning of the collection of data, in terms of the design of surveys and experiments.

There are two main philosophical approaches to statistics. The first is often referred as the Frequentist or classical approach. Procedures are developed by looking at how they perform over all possible random samples. The probabilities don't relate to the particular random sample that was obtained. In many ways this indirect method places the "cart before the horse". The alternative approach is Bayesian approach.

Bayesian statistics is a system for describing epistemological uncertainty using the mathematical language of probability. In the 'Bayesian paradigm,' degrees of belief in states of nature are specified; these are non-negative, and the total belief in all states of nature is fixed to be one. Bayesian statistical methods start with existing 'prior' beliefs, and update these using data to give 'posterior' beliefs, which may be used as the basis for inferential decisions. Bayesian methods are gaining popularity in main areas such as medical, marketing, cost effectiveness of medicines, terrestrial carbon dynamics, auditing, radiocarbon dating, setting water quality standards, food production, food technology, clinical trials and other fields where prediction and decision-making must follow for statistical analysis.

1.2 Objectives

The main objectives of our study are:

- Bayesian analysis of binary logistic regression model
- Bayesian analysis of binary logistic regression model without intercept
- Bayesian analysis of binary logistic regression model with intercept
- Bayesian analysis of binary logistic regression model with two explanatory variables
- To compute the posterior estimates for the different model with different priors
- Testing the hypotheses related to the parameters through Bayesian approach
- Comparison of Bayesian and Frequentist results

To achieve the objectives we went through different stages of research, its explanation is presented here:

Chapter 2 is basically concerned with the basic elements of Bayesian statistics which are considered to be the foundation of Bayesian analysis. The introduction of Bayesian technique and the idea of Bayesian econometrics are given along with the different kinds of prior distributions for the unknown parameters. The informative prior (Normal), noninformative priors i.e. Jeffreys prior, Haldane prior and uniform prior are explained. We have also discussed about likelihood function formulation and posterior distribution that is based on prior distribution and likelihood function. Bayesian hypotheses testing its similarities with the Frequentists testing and its advantages are presented. Advantages and disadvantages of Bayesian statistics are also presented.

Chapter 3 provides the explanation about Bayesian analysis for logistic regression model, the formulation of posterior distribution of parameters for these types of model. The concept of odds ratio is explained. Detailed review of the existing literature on the Bayesian logistic regression inferences is presented. Few of them are Croweder & Sweeting (1989), Zellner (1983), Munkin & Trivedi (2008), Poirier (1994), Choi et al., (2008), Tektas & Gunay (2008), Roman & Richard (2009), Frank Rijman (2008) and Bermudez et. al., (2007).

In chapter 4, we present the Bayesian analysis of logistic regression model without intercept under informative and noninformative priors. Data sets that we have used through out our study are presented in Table 4.1 with one explanatory variable. The derivation and introduction of noninformative priors are given and the complete steps for derivation of posterior distribution and the differentiation of posterior distribution for posterior modes are also given. Then for informative prior the range of hyperparameters is given and selected the appropriate. The idea of selecting hyperparameters are taken from Bian (1997), they assume the mean as zero for prior and check the posterior at different values of variance, as it is not a good practice to assume mean for prior as zero when selecting Normal as prior so, we have selected different values for both parameters mean and variance to check the posterior and select the appropriate values for hyperparameters. The posterior results and testing for the significance of these parameters are also presented. At the end, the classical results are compared with Bayesian estimates.

In chapter 5, we present the Bayesian analysis of logistic regression model with intercept under informative and noninformative priors using data sets given in chapter 4, Table 4.1. The complete process for finding posterior distribution with informative and noninformative priors is given. Along with the differentiation of these joint posteriors for posterior modes is also derived. Then for the informative priors the range of values of hyperparameters is given and selected the hyperparameters with minimum standard error. The hypotheses testing for the significance of the parameters is also done. Classical results and their comparison with Bayesian results are also given.

In chapter 6, we present the Bayesian analysis of logistic regression model with two explanatory variables under informative and noninformative priors using the data set given in Table 6.1. The complete process to find the posterior distribution for informative and noninformative priors along with the differentiation of these joint posterior for posterior modes is also presented. The range of hyperparameters is also given and selected the hyperparameters with minimum standard error. Then we present the hypothesis testing for the significance of parameters. The classical estimates and their comparison with Bayesian estimates are also provided.

In chapter 7, we have interpreted the results obtained by analysis in the previous chapters and the direction for further research is suggested. The programs executed for different calculations are given in Appendix. Books and numerous journals consulted during the research are listed in references.

Chapter 2

2.1 Introduction

This chapter is basically concerned with the basic terminologies of Bayesian Statistics which are considered to be the foundation of Bayesian analysis and we will also discuss the use of Bayesian techniques in Econometrics.

Section 2: Bayesian statistics is briefly discussed, Section 3: why we study Bayesian statistics. Section 4: describes the prior distribution such as Noninformative prior and Informative prior. Section 5: we discuss about the likelihood function. Section 6: Posterior distribution is defined. Sections 7, 8 & 9: Bayesian hypothesis testing, its similarities with the Frequentist testing and its advantages are also presented. Section 11: In this section we described the Advantages and Disadvantages of Bayesian Statistics and in Section 12: the difference between Frequentist and Bayesian Statistical Methods is given.

2.2 The Bayesian Statistics

Science inquiry is an iterative process of integrating accumulating information. Investigators assess the current state of knowledge regarding the issue of interest, gather new data to address remaining questions, and then update and refine their understanding to incorporate both new and old data. Bayesian inference provides a logical, quantitative framework for this process. It has been applied in a multitude of scientific, technological, and policy settings.

"Bayesian" refers to the Reverend Thomas Bayes. The development of probability theory in the early 18th century arose to answer questions in gambling, and to underpin the new and related ideas of insurance. A problem arose, known as the question of inverse probability: the mathematicians of the time knew how to find the probability that, say, 4 people aged 50 die in a given year out of a sample of 60 if the probability of any one of them dying was known. But they did not know how to find the probability of one 50-year old dying based on the observation that 4 had died out of 60. The answer was found by Thomas Bayes, and was published in 1763 (the year after his death). Like many educated men of his time, Bayes was both a clergyman and an amateur scientist/mathematician. His solution, known as Bayes theorem, underlies, and gave its name as, the modern Bayesian approach to the analysis of all kinds of data.

What we know as Bayesian statistics has not had a clear run since 1763. Although Bayes method was enthusiastically taken up by Laplace and other leading probabilists of the days, it fell into disrepute in the 19th century because they did not yet know how to handle prior probabilities properly. The first half of the 20th century saw the development of a completely different theory, now called Frequentist statistics. But the flame of Bayesian thinking was kept alive by a few thinkers such as Bruno de Finetti in Italy and Harold Jeffreys in England. The modern Bayesian movement began in the second half of the 20th century, spearheaded by Jimmy Savage in USA and Dennis Lindley in England, but Bayesian inference remained extremely difficult to implement until the late 1980s and early 1990s when powerful computers became widely accessible and new computational methods were developed. The subsequent explosion of interest in Bayesian statistics has led not only to extensive research in Bayesian methodology but also to the use of Bayesian methods to address pressing questions in diverse application areas such as astrophysics, weather forecasting, health care policy, and criminal justice.

Bayesian inference is an approach to a statistics in which all forms of uncertainty are expressed in terms of probability.

A Bayesian approach to problem starts with the formulation of a model that we hope is adequate to describe the situation of interest. We then formulated a prior distribution over the unknown parameters of the model, which is meant to capture our beliefs about the situation before seeing the data. After observing some data, we apply Bayes Rule to obtain a Posterior distribution for these unknowns, which take account of both the prior and the data. From this posterior distribution we can compute predictive distributions for future observations.

This theoretically simple process can be justified as the proper approach to uncertain inference by various arguments involving consistency with clear principles of rationality. Despite this, many people are uncomfortable with the Bayesian approach, often because they view the selection of a prior as being arbitrary and subjective. It is indeed subjective, but for this very reason it is not arbitrary. There is (in theory) prior beliefs. In contrast, other statistical methods are truly arbitrary, in that there are usually many methods that are equally good according to non-Bayesian criteria of goodness, with no principled way of choosing between them.

2.3 Why we study Bayesian?

There are certain reasons for which Bayesian approach is considered to be the better approach then the Classical approach.

- Bayesian statistics is preferred over Classical (Frequentist) Statistics because it is very useful in the situations where uncertainty is unavoidable.
- Parameter estimates along with confidence intervals or highest density region are calculated directly from the posterior distribution.

- Bayesian statistics is used for the prediction of future observations, which can be easily determine on the conditional probability distribution of the next observations given the sample data.
- Inference problems concerning parameter can easily be dealt with using Bayesian analysis.

2.4 Bayesian Econometrics

Bayesian econometrics is a branch of econometrics which applies Bayesian principles to economic modeling. The Bayesian principle is based on Bayes Theorem which states that the probability of "B" conditional on "A" is the ratio of joint probability of "A" and "B" divided by probability of "B". Bayesian econometricians assume that coefficients in the model have prior distributions. This approach was first propagated by Arnold Zellner (1983). He is known for his pioneering work in the field of Bayesian analysis and econometric modeling. In Bayesian analysis, he not only provided many applications of it but also a new information theoretic derivation of Bayes' theorem and generalizations of it that is 100% efficient information processing rules. As regards econometric modeling, he, in association with Franz Palm, developed the structural econometric, time series analysis approach for constructing new models and for checking the adequacy of old models that has been widely applied. In addition, he has been involved in many important applied econometric and statistical studies.

2.5 Prior Distributions

Scientific hypothesis typically are expressed through probability distributions for observable scientific data. These probability distributions depend on unknown quantities called parameters. In the Bayesian paradigm, current knowledge about the model parameters is expressed by placing a probability distribution on the parameters, called the "prior distribution". Often written as $p(\beta)$.

Also a "prior distribution" is a marginal probability, interpreted as a description of what is known about a variable in the absence of some evidence.

In Bayesian statistical inference, a prior probability distribution, often simply the prior, of an uncertain quantity p (i.e. suppose that p is the proportion of voters who will vote for the politician named Smith in a future election) is the probability distribution that would express one's uncertainty about p before the "data" (e.g., an opinion poll) are taken into account. It is meant to attribute uncertainty rather than randomness to the uncertain quantity. There are two types of priors: informative and Noninformative (or "reference"). Box and Tiao (1973) defined a Noninformative prior as one that provides little information relative to the experiment in this case the stock assessment data. Informative prior distributions, on the other hand, summarize the evidence about the parameters concerned from many sources and often have a considerable impact on the results.

2.5.1 Choice of Prior Distributions

A prior may be declared as an Achilles heel of Bayesian statistics, where the 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. If it varies from zero to one, we usually use Beta (Dirichlet) prior; for the range from zero to infinity, we select gamma prior, for minus infinity to infinity we usually use normal prior, etc. In the prior distribution, we quantify the uncertainty about the unknown parameter(s) in the form of a probability distribution, usually denoted by $p(\mu)$, and call it the prior distribution. In the Bayesian statistical inference, a prior probability distribution of an uncertain quantity μ is the probability distribution that would express one's uncertainty about μ, before the data set or evidence is taken into account. Since, the specification of prior is purely a subjective assessment of an expert; it makes the entire inference subjective in nature, which is the fundamental objection of rabid Frequentists to the Bayesian approach. Being subjective does not mean being non-scientific, as critics of Bayesian statistic often insinuate. On the contrary, vast amount of scientific information coming from theoretical and physical models is guiding in the specification of priors. Lindley's (2004) view is that `objectivity is merely subjectivity when nearly everyone agrees'. Such information is then merged with the data sets for better inference.

2.5.2 Noninformative Priors

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 noninformative priors, indifferent, ignorant, and vague or reference priors. Berger (1985) argues that Bayesian analysis, using noninformative priors, is the single most powerful method of statistical analysis in the sense of being the ad hoc method most likely to yield a sensible answer. The topic has an extensive literature, e.g., Jeffreys (1946, 1961), Bernardo (1979), Ghosh and Mukerjee (1992), Kass & Wasserman (1996) and Tibshirani (1989), propose the Bayesian analysis of unknown parameters using one of the most widely used noninformative priors, that is, a uniform (possibly improper) prior that routinely used by Laplace (1812).

Some attempts have been made at finding probability distributions in some sense, logically required by the nature of one's state of uncertainty; these are a subject of philosophical controversy. For example, (Jaynes 1968) has published an argument based on Lie groups that suggests that the prior for the proportion p of voters voting for a candidates, given no other information, should be the Haldane prior $p^{-1}(1-p)^{-1}$. If one is so uncertain about the value of the aforementioned proportion p that one knows only that at least one voter will vote for Smith and at least one will not, then the conditional probability distribution of p given this information alone is the uniform distribution on the interval [0, 1], which is obtained by applying Bayes theorem to the data set consisting of one vote for Smith and one vote against, using the above prior. The Haldane prior has been criticized on the grounds that it yields an improper posterior distribution that puts 100% of the probability content at either $p = 0$ or at $p = 1$ if a finite sample of voters all favor the same candidate, even though mathematically the posterior probability is simply not defined and thus we cannot even speak of a probability content.

A related idea, reference prior, was introduced by Bernardo (1979). Here, the idea is to maximize the expected Kullback-Leibler divergence of the posterior distribution relative to the prior. This maximizes the expected posterior information about X when the prior density is $p(x)$. The reference prior is defined in the asymptotic limit, i.e., one considers the limit of the priors so obtained as the number of data points goes to infinity. Reference priors are often the objective prior of choice in multivariate problems, since other rules (e.g., Jeffreys rule) may result in priors with problematic behavior.

The Jeffreys rule attempts to solve this problem by computing a prior which expresses the same belief no matter which metric is used. The Jeffreys prior for an unknown proportion p is $p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$, which differs from Jayne's recommendation.

Practical problems associated with Noninformative priors include the requirement that the posterior distribution be proper. The usual Noninformative priors on continuous, unbounded variables are improper. This need be a problem if the posterior distribution is proper. Another issue of importance is that if a Noninformative prior is to be used routinely, i.e., with many different data sets, it should have good Frequentist properties.

2.5.3 Informative Priors

An informative prior expresses specific, definite information about a variable. An example is a prior distribution for the temperature at noon tomorrow. A reasonable approach is to make the prior a normal distribution with expected value equal to today's noontime temperature, with variance equal to the day-to-day variance of atmospheric temperature.

This example has a property in common with many priors, namely, that the posterior from one problem (today's temperature) becomes the prior for another problem (tomorrow's temperature); pre-existing evidence which has already been taken into account is part of the prior and as more evidence accumulates the prior is determined largely by the evidence rather than any original assumption, provided that the original assumption admitted the possibility of what the evidence is suggesting. The terms "prior" and "posterior" are generally relative to a specific datum or observation.

The following two sections outline the two techniques used most frequently to develop informative prior distributions and the final section provides some advice on default choices for priors when applying typical methods of fisheries stock assessment.

(i) Expert Opinion

In principle, one of the most powerful methods for developing informative priors is to synthesize the information from a group of experts. For example, international Whaling Commission (1995) developed priors for the assessment of the Bering-Chukchi-Beaufort seas stock of bowhead whales by consensus. Although the development of priors by consensus risks all the problems related to the impact of the subjective biases of various parties in the assessment process (arguably priors developed using expert opinion are examples of "dreamt up" priors, to use an expression we used in the previous section), this approach can be successful. The members of the assessment group were provided with the values for other biological parameters (growth, natural mortality, etc.) for the entire assessment group tended to be more pessimistic than those suggested by the industry members; this was nevertheless generally regarded as a successful attempt at specifying a prior.

A potentially major problem with the development of priors by consensus is that different "experts" will suggest different priors. It is far from a trivial exercise (theoretically) to pool such priors to form a "consensus prior" (and it is impossible to include more than one prior for each parameter in a Bayesian assessment). Unfortunately, relatively little work has been directed recently at this problem. We recommend that the various priors be multiplied together and than normalized because at least this procedure has the desirable property that the assessment results are independent of whether the priors are pooled and than the assessment conducted or whether assessments conducted using each alternative prior in turn and the results then pooled. One very undesirable feature of this approach to pooling, however, is that if one expert believes that some parameter value/model has zero probability, the posterior is forced to be consistent with this opinion. Therefore, if this approach is to be used, our earlier advice that no plausible value for a parameter should be assigned zero probability should be followed.

(ii) Data Summaries/Meta-Analysis

If the parameters of the stock assessment model are chosen to be independent of the parameter that scales the population, data for other species and stocks can be used to construct priors for the species for which an assessment is needed. This approach to conducting priors is known as meta-analysis. Methods for constructing priors using data for other stocks and species range from simply tabulating the estimates to hierarchical meta-analysis. Simple tabulation methods can be extended by fitting a smooth functional form to the data and by weighting each estimate by a measure of its uncertainty and comparability to the stock and species for which an assessment is required. Hierarchical meta-analysis is a more formal method for developing a prior for a parameter from values for that parameter for other stocks under the assumption that the stocks differ in that parameter.

"Selection bias" is a potential problem when developing a prior using data for similar stocks and species. Assessments in the literature tend to be for large productive populations (small, less productive populations in general receiving less research funding). If the stocks considered are not representative of all similar stocks, an inappropriate prior may be selected.

2.6 Likelihood Function

Maximum-likelihood estimation was recommended, analyzed and vastly popularized by R. A. Fisher between 1912 and 1922 (although it had been used earlier F. Y. Edge worth). Reviews of the development of maximum likelihood have been provided by a number of authors. Maximum likelihood estimation (MLE) is a popular statistical method used for fitting a statistical model to data, and providing estimates for the model's parameters. The method of maximum likelihood corresponds too many well-known estimation methods in statistics. For example, suppose you are interested in the heights of adult female giraffes. You have a sample of some number of adult female giraffes, but not the entire population, and record their heights. Further, if we are willing to assume that heights are normally distributed with some unknown mean and variance. The sample mean is then the maximum likelihood estimator of the population mean, and the sample variance is a close approximation to the maximum likelihood estimator of the population variance. For a fixed set of data and underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them. Maximum likelihood estimation gives a unique and easy way to find a solution in the case of the normal distribution and many other problems, although in very complex problems this may not be the case. If a uniform prior distribution is assumed over the parameters, the maximum likelihood estimate coincides with the most probable values.

Suppose there is a sample $x_1, x_2, \ldots, x_n$ of n i.i.d observations, coming from an unknown distribution $f_0(\cdot)$. It is however known that the function f_0 belongs to a certain family of distributions $\{f(\cdot|\theta), \theta \in \Theta\}$, called the parametric model, so that $f_0 = f(\cdot|\theta_0)$. The value θ_0 is unknown and is referred to as the "*true value*" of the parameter. It is desirable to find some $\hat{\theta}$ (the *estimator*) which would be as close to the true value θ_0 as possible. Both the observed variables x_i and the parameter θ can be vectors. The variables x_i may be non-iid, in which case the formula below for joint density will not separate into individual terms; however the general principles would still apply. To use the method of maximum likelihood, one first specifies the joint density function for all observations. For iid sample this joint density function will be

$$f(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdots f(x_n|\theta) \tag{2.1}$$

In basic statistics and in many other problems, we may extend the domain of the density function so that the density is also a function of the parameter θ. Then, for a given sample of data with observed values x_1, x_2, ..., x_n, the extended density can be considered a function of the parameter θ. This extended density is the likelihood function of the parameter:

$$\mathcal{L}(\theta \mid x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i|\theta). \tag{2.2}$$

However, in general, the likelihood function is not a probability density. In fact, it need not be an additive function, so it is not a probability measure. In practice it is often more convenient to work with the logarithm of the likelihood function, $\ln L$, called the log-likelihood, or its scaled version, called the *average log-likelihood*:

$$\ln \mathcal{L}(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i|\theta), \qquad \hat{\ell} = \frac{1}{n} \ln \mathcal{L}. \tag{2.3}$$

The hat over ℓ indicates that it is akin to some estimator. Indeed, $\hat{\ell}$estimates the expected log-likelihood of a single observation in the model. The method of maximum likelihood estimates θ_0 by finding a value of θ that maximizes$\hat{\ell}(\theta|x)$. This method of estimation is a maximum likelihood estimator (MLE) of θ_0:

$$\hat{\theta}_{\text{mle}} = \underset{\theta \in \Theta}{\arg\max}\, \hat{\ell}(\theta \mid x_1, \ldots, x_n). \tag{2.4}$$

A MLE estimate is the same regardless of whether we maximize the likelihood or the log-likelihood function.

2.7 Posterior Distribution

The notion of a posterior distribution comes from Bayesian statistics. Under the Bayesian approaches, prior beliefs about parameters are combined with sample information to create updated or posterior beliefs about the parameters. In the case of empirical Bayes estimators, the prior information comes from the sample data as well.

The posterior information is proportional to the product of the prior information and the sample information.

The posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence is taken into account. The posterior probability distribution of one random variable given the value of another can be calculated with Bayes theorem by multiplying the prior probability distribution by the likelihood function, and then dividing by the normalizing constant, as follows:

$$p(\beta/X) = \frac{L(X/\beta)P(\beta)}{\int L(X/\beta)P(\beta)d\beta}, \tag{2.5}$$

Gives the posterior probability density function for a random variable β (parameter) given the data $X = x$,

Where

$P(\beta)$ is the prior density of β

$L(X/\beta)$ is the likelihood function as a function of x

$\int L(X/\beta)P(\beta)d\beta$ is the normalizing constant, and

$p(\beta / X)$ is the posterior density of β given the data $X = x$.

2.8 Bayesian Hypothesis Testing

A statistical hypothesis test is a method of making statistical decisions using experimental data. In statistics, a result is called statistically significant if it is unlikely to have occurred by chance. Hypothesis testing is sometimes called confirmatory data analysis, in contrast to exploratory data analysis. In frequency probability, these decisions are almost always made using null-hypothesis tests (i.e., tests that answer the question *Assuming that the null hypothesis is true, what is the probability of observing a value for the test statistic that is at least as extreme as the value that was actually observed?*). One use of hypothesis testing is deciding whether experimental results contain enough information to cast doubt on conventional wisdom. Bayesian hypothesis testing is less formal than .non-Bayesian varieties. In fact, Bayesian researchers typically summarize the posterior distribution without applying the rigid decision process. Since social scientists don't actually make important decisions based on their findings, posterior summaries are more than adequate. If one wanted to apply a formal process, Bayesian decision theory is the way to go because it is possible to get a probability distribution over the parameter space and one can make the expected utility calculations based on the costs and benefits of different outcomes. Since in Bayesian analysis, the task of deciding between H_0 and H_1 is conceptually more straightforward. One merely calculates the posterior probabilities $\alpha_0 = P(\Theta_0 \mid x)$ and $\alpha_1 = P(\Theta_1 \mid x)$ and decides between H_0 and H_1 accordingly. The conceptual advantages are that α_0 and α_1 are actual (subjective) probabilities of the hypothesis in light of the data and prior opinions.

2.9 Similarities between Bayesian and Frequentist Hypothesis Testing

(i) Maximum likelihood estimates of parameter means and standard errors and Bayesian estimates with flat priors are equivalent.

(ii) Asymptotically, the data will overwhelm the choice of prior, so if we had infinite data sets, priors would be irrelevant and Bayesian and Frequentist results would converge.

(iii) Frequentist one-tailed tests are basically equivalent to what a Bayesian would get using credible intervals.

The most important pragmatic difference between Bayesian and Frequentist hypothesis testing is that Bayesian methods are poorly suited for two-tailed tests. Because the probability of zero in continuous distribution is zero. The best solution proposed so far is to calculate the probability that, say a regression coefficient is in some range near zero, e.g. two sided p-value = pr(-e<B<e).

However, the choice of 'e' seems very adhoc unless there is some decision theoretic basis. The other important difference is more philosophical. Frequentist p-values violate the likelihood principle.

2.10 Advantages of Bayesian Testing

(i) A defaults formula exists for all situations:

$$pr(H_0 \mid data) = \left[1 + \frac{\int\int f(x;\theta) f(x^*;\theta) f(x^*;\theta_0) dx^* d\theta}{\int f(x;\theta_0) f(x^*;\theta) d\theta} \right]^{(-1)} \tag{2.6}$$

Where x^* is independent (unobserved) data of the smallest size such that the above integral exists?

(ii) Posterior probabilities allow for incorporation of personal opinion, if desired. Indeed, if the published default posterior probability of H_0 is P^*, and the prior probability of H_0 is P_0, then the posterior probability of H_0 is:

$$pr(H_0 \mid data) = \left[1 + \left(\frac{1}{P_0} - 1\right)\left(\frac{1}{P^*} - 1\right)\right]^{(-1)} \qquad (2.7)$$

Example: In binomial say $P^* = 0.52$.

A "skeptic "has $P_0 = 0.1$; hence $pr(H_0 \mid data) = 0.11$.

A "believer" has $P_0 = 0.9$; hence $pr(H_0 \mid data) = 0.91$.

(iii) Posterior probabilities are not affected by the reason for stopping experimentation, and hence do not require rigid experimental designs (as do classical testing measures).

(iv) Posterior probabilities can be used for multiple models or hypothesis.

2.11 Advantages and Disadvantages of Bayesian Statistics.

Here we will first considered the advantages of Bayesian statistics due to which the branch of statistics has a valuable respect among the class of statistician known as Bayesian statistician. Following are the advantages of Bayesian statistics.

(i) Exact inferences (e.g., confidence interval) which do not rely on large sample approximations, are available through Bayesian approach.

(ii) Bayesian answers have simple interpretation: "let 95% Bayesian interval for θ is (0.25, 0.87)" mean "there is probability 0.95 that θ is between 0.25 and 0.87". Interpretation of Frequentist interval is hard, and most users tend to falsely interpret them as in the above.

(iii) Interpretation of Bayesian interval depends on the data at hand, but not so for Frequentist intervals. This can cause logical (or coherency) problems.

(iv) Elimination of nuisance parameters is conceptually straightforward, and is also easy due to advances in Bayesian computing. This convenience is a result of Bayesian analysis being a logically simple and easy approach.

(v) Stopping rules are irrelevant in Bayesian analysis. This makes Bayesian analysis much easier to use in areas such as clinical trials. In clinical trails, experimenters would like to analyze the data frequently, and make decision without having to adhere to a pre-specified design protocol. Bayesian analysis allows this. But, such flexibility is very difficult to achieve using Frequentist methods.

(vi) Bayesian approach allows flexibility of models. Highly complex models (with many structures) can be fitted. This is making the Bayesian approach more appealing in many areas.

(vii) Bayesian learning methods interpolate all the way to pure engineering. When faced with any learning problem, there is a choice of how much time and effort a human vs. a computer puts in. (For example, the mars rover path finding algorithms are almost entirely engineered.) When creating an engineered system, you build a model of the world and then find a good controller in that model. Bayesian methods interpolate to this extreme because the Bayesian prior can be a delta function on one model of the world. What this means is that a recipe of "think harder" (about specifying a prior over world models) and "compute harder" (to calculate a posterior) will eventually succeed. Many other machine learning approaches don't have this guarantee.

(viii) Bayesian and near-Bayesian methods have an associated language for specifying priors and posteriors. This is significantly helpful when working on the "think harder" part of a solution.

(ix) Bayesian learning involves specifying a prior and integration, two activities which seem to be universally useful.

Now we consider the disadvantages of Bayesian approach that remain a vital cause for not being used extensively.

(i) It requires us to specify a prior distribution for all parameters. When there is concrete prior knowledge about the parameters, it can be done, and should be done. But, in many cases, prior knowledge is either vague, or non-existent, and that makes it very difficult to specify a unique prior distribution. Different opinion, may suggest different priors, and arrive at different answers. Question of "objectivity is concern here.

- In practice, researcher often overcome this by using certain non-informative or default priors. These are priors that are easy to specify and hold little or no prior information about the parameters.
- When there is sufficient data (large sample), prior do not affect the answer (likelihood will dominate), and so the answer will be the same, regardless of what prior is used.
- "Reality" is there an objective answer?
- Scientist often disagrees on the conclusion and interpretation of results that are different due to the different prior information used.

(ii) Bayesian methods typically involve high-dimensional integrals. If the statistical problem involves four parameters (e.g., comparing two normal means), then the inference involve 4-dimensional integration. No longer a serious concern after the advent of Markov Chain Monte Carlo (MCMC) methods. However, MCMC can be time consuming in complex problems. But, often it is worth the effort, as Bayesian methods allows fitting complex models without resorting to large sample approximation.

(iii) It turns out that specifying a prior is extremely difficult. Roughly speaking, we must specify a real number for every setting of the world model parameters. Many people well-versed in Bayesian learning don't notice this difficulty for two reasons:

- They know languages allowing more compact specification of priors. Acquiring this knowledge takes some significant effort.
- They lie. They don't specify their actual prior, but rather one which is convenient. (This shouldn't be taken too badly, because it often works.)

(iv) Let's suppose I could accurately specify a prior over every air molecule in a room. Even then, computing a posterior may be extremely difficult. This difficulty implies that computational approximation is required.

(v) The "think harder" part of the Bayesian research program is (in some sense) a "Bayesian employment" act. It guarantees that as long as new learning problems exist, there will be a need for Bayesian engineers to solve them.

Chapter 3

Bayesian Logistic Regression & Literature Review

3.1 Introduction

In this chapter we will discuss briefly the Bayesian logistic regression analysis, their detail results will be presented in next chapters. We will discuss the basic technique for analyzing binary logistic regression model with Bayesian approach in section 2. In section 3 the literature review is given.

3.2 Logistic Regression

Logistic regression is used by practitioners and researchers in many fields, but is undoubtedly used most frequently in medical and biomedical applications. Maximum likelihood is generally the estimation method of choice. A check of the Science Citation Index reveals that 2770 papers were published in 1999 in which "logistic regression" appeared in either the title or among the key words (king & Ryan 2002). Since in many fields of application, dichotomous qualitative models have been studied using non-Bayesian techniques. For example: Amemiya (1981), Hausman and McFadden (1984) and McFadden (1981). However, recently there has been great interest in Bayesian analysis of dichotomous and polychotomous response models. This can be seen in McCulloch et. al. (1999), Albert & Chib (1993), Koop & Poirier (1993), Stukel (1998), Basu & Mukhopadhayay (2000) and Bazan et. al. (2006).

3.2.1 Logistic Regression Model

The logistic regression model is perhaps the most widely used among researchers whose goal is to model binary dependent variables. The first type of discrete variable addressed is probably the most common: a binary or dichotomous dependent variable. It is unwise to use Ordinary Least Square (OLS) when confronted with a binary dependent variable. So the alternative regression models are implemented to handle this difficulty.

Logistic regression is a form of statistical modeling that is often appropriate for categorical outcome variables. It describes the relationship between a categorical response variable and a set of explanatory variables. The response variable is usually Dichotomous, but it may be polychotomous, that is, have more than two response levels. These multiple-level response variables can be nominally or ordinally scaled. Here our research interest is Dichotomous response; typically the two outcomes are yes and no.

Now let us suppose we are interested in explaining the distribution of some dependent variable, yet it has only two possible outcomes. For example, this dependent variable might measure whether or not respondents in a sample support the death penalty, whether or not respondents graduated from college. In each of this examples, the variable is often coded as [0, 1], with 0 indicating "no" and 1 indicating "yes". The main difficulty for a regression model occurs when the researcher wishes to use a binary variable as the dependent variable. It should be clear that this variable does not and will not follow a normal or Gaussian distribution. Rather, it is distributed as a binomial random variable. But if a researcher still want to predict this variable within a regression-like context, then Logistic Regression Model may be a suitable choice. The key to this model is that, rather than modeling the dependent variable directly (i.e. estimating the expected value of the dependent variable "Y" for some combination of independent variables), we estimate the probability that Y=1. Just like in linear regression we assume that some set of "X" variables is useful for predicting the Y values, but we are claiming that this set predicts the probability that Y=1 (assuming we have coded the dependent variable as [0, 1]. This transformation from directly modeling the dependent variable to modeling some variation of it is only possible with the help of a link function.

Sometimes the term "logistic regression" is restricted to analyses that include continuous explanatory variables, and the term "logistic analysis" is used for those situations where all the explanatory variables are categorical. Here we will focus on logistic regression only.

The basic formula for estimating Y=1 consists of transforming the regression equation as.

$$P(Y=1|X_1,X_2,...,X_k)=\frac{1}{1+\exp(-\beta_o-\beta_1X_1-\beta_2X_2-...-\beta_kX_k)} \quad (3.1)$$

Then

$$P(Y=0)=1-P(Y=1)=\frac{1}{1+\exp(\beta_o+\beta_1X_1+\beta_2X_2+...+\beta_kX_k)} \quad (3.2)$$

The part of the denominator in parentheses should remind us of the standard linear regression model. But note that in this function it is transformed in what seems to be an unusual way. This part is multiplied by -1, exponentiated, added to 1 and then inverted. The whole function is called the Logistic Function (Hoffmann 2004). Another form of this equation that is often used is:

$$P(Y=1|X_1,X_2,...,X_k)=\frac{\exp(\beta_o+\beta_1X_1+\beta_2X_2+...+\beta_kX_k)}{1+\exp(\beta_o+\beta_1X_1+\beta_2X_2+...+\beta_kX_k)} \quad (3.3)$$

The quantity β_0 is the intercept parameter; the X's are the k explanatory variables and $\beta's$ are the regression parameters:

We can write above in odds form as.

$$\frac{P(Y=1|X_1,X_2,...,X_k)}{1-P(Y=1|X_1,X_2,...,X_k)}=\exp(\beta_o+\beta_1X_1+\beta_2X_2+...+\beta_kX_k) \quad (3.4)$$

By taking natural logarithms on both sides, you obtain a linear model for the Logit:

$$Log\left(\frac{P(Y=1|X_1,X_{2,...,}X_k)}{1-P(Y=1|X_1,X_{2,...,}X_k)}\right)=\beta_o+\beta_1X_1+\beta_2X_2+...+\beta_kX_k \qquad (3.5)$$

The Logit is the log of an odd. The log odds for kth group can be written as the sum of an intercept and a linear combination of explanatory variable values multiplied by the appropriate parameter values.

$$Log\left(\frac{p(Y=1|X_1,X_{2,...,}X_k)}{1-p(Y=1|X_1,X_{2,...,}X_k)}\right)=\beta_o+\sum_{i=1}^{k}\beta_iX_i \qquad (3.6)$$

This result allows you to obtained the model- predicted odds ratio for variation in the X's by exponentiating model parameter estimates for the $\beta's$.

Besides taking the familiar linear form, the logistic model has the useful property that all possible values of $\beta_o+\sum_{i=1}^{k}\beta_iX_i$ in $(-\infty,\infty)$ map into $(0,1)$ for $p(Y=1)$. Thus, predicted probabilities produced by this model are constrained to lie between 0 and 1. This model produces no negative predicted probabilities and no predicted probabilities greater than 1. Maximum likelihood methods are generally used to estimate $\beta's$.

Logistic regression has applications in the fields such as epidemiology, medical research, banking, market research, and social research. One of its advantages is that model interpretation is possible through odds ratio, which are functions of model parameters.

3.2.2 Odds Ratio

The odds ratio is a measure of effect size, describing the strength of association or non-independence between two binary data values. It is used as a descriptive statistic, and plays an important role in logistic regression. Unlike other measures of association for paired binary data such as the relative risk, the odds ratio treats the two variables being compared symmetrically, and can be estimated using some types of non-random samples.

The definition of odds ratio in terms of group wise odds can be presented as: The odds ratio is the ratio of the odds of an event occurring in one group to the odds of it occurring in another group, or to a sample-based estimate of that ratio. These groups might be men and women, an experimental group and a control group, or any other dichotomous classification. If the probabilities of the event in each of the groups are p_1 (first group) and p_2 (second group), then the odds ratio is:

$$\frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{p_1/q_1}{p_2/q_2} = \frac{p_1 q_2}{p_2 q_1}, \tag{3.7}$$

Where $q = 1 - p$. An odds ratio of 1 indicates that the condition or event under study is equally likely to occur in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely to occur in the first group, and an odds ratio less than 1 indicates that the condition or event is less likely to occur in the first group. The odds ratio must be greater than or equal to zero if it is defined. It is undefined if $p_2 q_1$ equals zero. The odds ratio is used extensively in the healthcare literature. The odds ratio may be a misleading approximation to relative risk if the event rate is high (Deeks (1996) and Davies et al. (1998)). Since the odds ratio is difficult to interpret, why is it so widely used? First, odds ratio can be calculated for case-control studies whilst relative risks are not available for such studies. Second, if we use an analysis method that corrects for confounding factors, such as logistic regression, this will report results as odds ratio.

3.2.3 Bayesian Logistic Regression Analysis

Since we know that while using maximum likelihood method (MLE) for the estimation of regression coefficients it may mislead when we have small sample data sets as it happened in the field of medical science, because MLEs are usually based on asymptotic theory. Griffiths et. al. (1987) found that MLEs have significant bias for small samples. But this problem can be handled by using Bayesian technique while estimating regression parameters.

Let us considered here that the response variable y_i is categorical in nature with binary options coded as [0, 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 $n \leq N$ observations. Then for a sample of n observations the likelihood function is:

$$L(\beta \mid data) = \prod_{i=1}^{n} \left\{ p_i^{y_i} (1-p_i)^{1-y_i} \right\} \tag{3.8}$$

In the dichotomous response models $p_i = H(x_i'\beta)$, where $x_i = (x_{i1}, x_{i2}, ..., x_{ik})'$ is a $k \times 1$ vector of covariates, and $\beta = (\beta_1, \beta_2, ..., \beta_k)'$ is a $k \times 1$ vector of regression coefficients. Then the likelihood function can be written as:

$$L(\beta \mid data) = \prod_{i=1}^{n} \left\{ H(x_i'\beta)^{y_i} (1 - H(x_i'\beta))^{1-y_i} \right\} \tag{3.9}$$

Now the Bayesian analysis for the logistic model can follow the usual pattern for all Bayesian analysis i.e.

(i) Write down the likelihood function of the data as given in equation (3.9).

(ii) Assume a prior distribution over all unknown parameters.

(iii) Use Bayes theorem to find the Posterior distribution over all parameters.

Now for prior distribution in general, any prior distribution can be used, depending on the available prior information. The choice can include informative prior distributions if something is known about the likely values of the unknown parameters, or "diffuse" or "noninformative" priors if either little is known about the coefficient values or if one wishes to see that the data themselves provide as inferences.

Now let us consider the prior for the unknown regression coefficients as $p(\beta)$ then the posterior distribution for β is given as:

$$p(\beta \mid data) \propto L(\beta \mid data) \times p(\beta) \qquad (3.10)$$

Of course, the above expression has no closed form expression and even if it did, we would have to perform multiple integration to obtain the marginal distribution for each regression coefficient. So to solve the above function, SAS package help us a lot for the numerical solution of above function. As was the for Frequentist inference, taking exp(β) provides the odds ratio for a one unit change of that parameter.

3.3 Literature Review

Cengiz et al., (2001) illustrates how to model the binary logistic regression by using Bayesian approach. Binary response data is modeled using Binomial Distribution while the binary data have a Bernoulli distribution. The objective is to improve the accuracy and predictions and decision making by investigating logistic regression model in specific context of assessing Erythrocyte Sedimentation Rate (ESR). So, for this purpose the Author investigates by using and analyzing the five cases in which they present suitable priors distribution. When there is little prior information available, in these circumstances a vague prior is used. The standard choice is to use invariant prior proposed by Jeffrey's. They also use uniform and improper prior and then compare these results with classical inferences.

El-Sayyad (1973) concentrates upon a problem that the presences of any type of trend in the means of Poisson distribution while it change exponentially. Since, in simple classical method it is observed by testing the parameter 'β' of the Poisson model. The Bayesian approach is introduced here and the exact Bayesian distribution of ' β ' is derived and the Bayesian approximation is suggested which prove to be very useful. Then by using three method's i.e. Classical, Bayesian and Bayesian approximation with the help of an example the results are obtained and compared which concludes that Bayesian approach provide a better approximation then Classical.

Croweder & Sweeting (1989) viewed in context of an investigation conducted in the department of microbiology at Surrey University in which Fungal Spores are introduced into the earth surrounding the root of a plant. But with concern to a particular Question "whether the final alignment of the tube tip is random" a Bivariate case of Binomial distribution is studied. The sample information about parameter 'p' comes from Marginal distribution alone and the information about 'q' comes from Conditional distribution alone which are drive from Bivariate Binomial distribution. To study the behavior of the posterior parameter the sample size is increased and as the sample size increases the posterior parameter is approximately independent. This shows that Bayesian Conjugate Prior distribution arises from prior independence.

Zellner (1983) illustrates the usefulness of Bayesian approach by considering the different problems of Econometric models and shows that the Bayesian results are more appropriate then the usual Classical technique. He consider the problem of hypothesis analysis in which he point out the doubtful choice of significance level in which usually no attention is given to power consideration, but it seems highly probable that Bayesian analysis of hypothesis would yield more satisfactory results. He also consider the case of reference informative prior (RIP) as it is difficult to asses the prior for regression coefficient especially in Logistic Regression. He suggests that if we use Jeffrey or Jeffrey like prior i.e. RIP it will lead to a simple result. At the end he considers the prediction problem and concludes that prediction with the inclusion of prior knowledge is satisfactory then the usual Classical approach. Finally for complicated likelihood functions it is mention that numerical integration techniques are very helpful in analyzing posterior probability density functions and checking the validity of asymptotic and other approximations techniques.

Munkin & Trivedi (2008) develop an estimation procedure for the Ordered Probit Model with endogenous covariates by using the Bayesian approach and name it as the Ordered Probit Model with Endogenous Selection (OPES). They analyze the effect of endogenous dependent variable i.e. strongly agree, agree, disagree, strongly disagree etc. they model the endogeniety using a correlated latent variable structure. Then Markov Chain Monte Carlo (MCMC) method is used to approximate the posterior distribution of the parameters and treatment effect. This study is applied by analyzing the effects of different types of medical assurance plans on the level of hospital care utilization by the USA adult population and in their illustration they find the evidence that controlling for endogeniety is important.

Poirier (1994) uses the Jeffreys' prior for Logit models with covariates. He compares the properties of Jeffreys' prior with other priors that are mostly used for Logit models. Like natural conjugate priors, normal priors etc and he shows that Jeffreys' prior is not recommended in Conditional Logit models and it act like a neutral natural conjugate prior in Multinomial Logit models. The case of Jeffreys' prior with covariates has a substantial impact on its interpretation and three of which are discussed. At the end it is illustrated that Jeffreys' prior in context of Logit models and in the case of simple multinomial Logit models its properties with no covariates offers a little guidance for the cases involving covariates.

Albert & Chib (1993) illustrate that the categorical response regression model in Classical approach is fit by maximum likelihood method (MLE), but this approach is questionable when sample size is small as MLE is purely a large sample theory. So in this situation satisfactory results can be obtained by using Bayesian approach as classical approach cannot provide satisfactory results. They use Probit model for Binary outcomes as with the inclusion of latent variable it follows the structure of normal distribution and the value of latent variable is simulated by using truncated normal distribution. Then Gibbs sampling is used for posterior parameter estimation. So the Probit model on the binary response is connected with normal linear model on the continuous latent data response as we know that Probit model use the Cumulative density function (cdf) of normal. The exact binary analysis is performed and the result proves that it is better then usual MLE. At the end the case is also extended for Multinomial Logistic regression.

Choi et al., (2008) take a study of Bernoulli trials and estimate the parameters of modeled relationship between the covariates and the success probabilities that are based on Bayesian perspective by using the Markov Chain Monte Carlo (MCMC) algorithm on the available data. This study is also applied on real data. So a method is set with the help of above technique to estimate the parameters of the Logistic regression model when individual observations are missing but the aggregate information are available with covariate values. While using MCMC technique the missing observations are also considered as additional parameter to be estimated. At the end the results are compared with usual Classical techniques that handle the missing values case like Expectation maximization algorithm and Error-in-variables regression technique and the results proves that in this particular case of missing observations the Bayesian approach provide better results then the technique introduced by Classical.

Tektas & Gunay (2008) illustrate the basic objective of analyzing the Probit and Logit models by using Bayesian techniques proposed by Albert and Chib. The results are compared with usual Classical approach. It is shown in this article that Classical approach does not provide satisfactory results when the sample size is small while the Bayesian approach is best and suitable choice for this situation of small sample size. The parameters are estimated by using Gibbs sampling and Data augmentation algorithm together. The data is augmented by adding a set of latent variables (Z) into the model as latent variable is a continuous variable so the Conditional distribution of parameters given latent variable is a normal distribution whose mean is easy to compute. So Gibbs sampling is then use to calculate the posterior distribution for the parameter 'β'. At the end Logit and Probit models are estimated by using Bayesian approach. So the obtained results by using Bayesian approach are compared with usual Classical methods like Ordinary least square (OLS) and MLE. So the resulted table shows that Bayesian approach is better then Classical as the Bayesian results are much improved.

Tanner & Wong (1987) present an iterative method for the computation of Posterior distributions. This method is used when the data can be augmented in such a way that it become easy to analyze the augmented data and it is easy to generate the augmented data given the parameter. Augmentation is done by using the Latent variable i.e. 'Z' that is unobserved. The Author presents the basic algorithm and illustration by giving example. After that he also applied this method for Multivariate Normal distribution with missing values. Then Dirichlet sampling procedure is used to approximate sampling for Posterior distribution in complex Models and this procedure is applied to social survey data using Log-Linear model, then at the end with help of same example Bayesian modeling is used and the results are compared.

Rossi (1996) works on the existence of Bayes estimators for the Binomial Logit Model. As it is known that on finite maximum of the likelihood may not exist for certain configuration of the data. The importance is made on that, under what conditions the Posterior will be proper and when Posterior moments exists before proceeding to make numerical approximations to these moments, when we have Dichotomous dependent variable, it is important to calculate Posterior means. The Posterior density is obtained by using the Diffuse Prior. At the end the sufficient condition for integral convergence is compared with the condition provided by Zellner & Rossi (1984).

Silvapulle (1981) discusses and attempt to estimate maximum likelihood estimators for logit models that were first arose in the analysis of relationship of Psychiatric "caseness" to scores on a Psychiatric screening questionnaire. General Health Questionnaire (GHQ) to 120 patients attending a general practitioners surgery and also give each one a standardized Psychiatric interview by classifying as case/non-case i.e.

$$Logit\left[p_i(case)\right] = \beta_0 + \beta_1 x$$

Where x = GHQ score and $Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$ $Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$ for the full set of data. So, the convex analyses are used for proper estimation of maximum likelihood estimators.

Chen & Ibrahim (2003) propose a novel class of conjugate priors for the family of generalized linear models they discuss the elicitation issues that may occur during the application of different techniques available for the elicitation of hyperparameters. They developed theorems characterizing the property and existence of moments of the priors under various setting, examine asymptotic property and relationship with normal, their approach is based on the ground of specifying a prior prediction y_0 for the response vector of the current study and a scalar precision parameter a_0 which quantifies ones prior belief in y_0. Then a conjugate prior is specified for β regression coefficient with the help (y_0, a_0) along with explanatory variables. They also study the generalized linear models with dispersion parameter at its different values and check the effect on prior for fixed a_0 and random a_0. Also the results are illustrated with the help of an example and in numerical results it was observed that as a_0 increases the prior and posterior estimation are closer to each other.

Gelman et. al., (2008) Propose a new prior distribution for classical logistic regression models constructed by first scaling all non binary variables to have mean zero and standard deviation 0.5, then place independent t-prior distribution with Cauchy distribution as default prior that has mean zero and standard deviation 2.5, then a logistic regression model is fitted by using these priors and with the help of EM algorithm into the usual iteratively weighted least square. This default prior is recommended for further study as it has the advantage of giving results even in the case of complete separation in logistic regression. This is useful for routine data analysis as well as chain equations for missing-data imputation in which each variable is modeled with missing data. Then the logistic regression estimates are computed including prior distribution by applying Gibbs sampling and Metropolis algorithms. These computations are made in R package by defining a new function "bayesglm" where approximated posterior mode and variance are computed and used for further analysis. Since, the results are computed by using Cauchy prior, t-distribution as prior and normal prior and the results are compared with classical (generalized linear model) glm results. It is observed the default prior that is independent Cauchy distribution for all logistic regression coefficients each centered at zero and with scale parameter 10 for intercept and 2.5 for all other coefficients, with posterior modes as a point of estimate can be a usual approach to be adopted.

Eaves & Chang (1992) proposed a posterior mode estimator, which arise from simply expressed prior opinion about expected outcomes, roughly as follows, a conjugate prior is used to obtain the posterior modes and its covariance by using the conventional maximum likelihood computations. Then within the family of conjugate prior a reference prior is proposed to obtain the inferences about the regression vector for linear design of the canonical link. A set of subjective prior upper and lower percentage points for the expected outcomes can be used to determine a conjugate family member. They use the Jeffreys prior obtained the posterior modes, reference prior is also used to obtained the posterior modes and variance function to obtained the estimates and at the end the results are compared with usual classical approach.

Groenwald & Mokgatlhe (2005) suggest a method for the simulation of samples from the exact posterior distribution of the parameters in logistic regression. This method is based on the principle of data augmentation and on the induction of latent variable. Since in Bayesian logistic regression all conditional distributions are intractable but with the introduction of latent variable all conditional distribution are uniform and the Gibbs sampling is easily applicable then they extend this technique and applied with nominal or ordinal polychotomous data. So, in section 3 of this paper they extend this technique for multiple response categories and in section 4 study for ordinal response with thresholds or cut off points are presented. In section 5 the data augmentation technique is applied to model selection via Bayes factors. The marginal likelihood under a particular model can be calculated by running additional Gibbs cycles, one for each parameter in the model. Then at the end this technique is illustrated by analysis two real life examples.

Bian (1997) presents Bayesian inferences for location parameter of a family of location-scaled distributions i.e. student-t and normal distribution. He develops Bayesian estimators for the location parameter of a location-scale distribution for this purpose they use modified maximum likelihood estimator (MMLE). As the Bayesian estimators are defined by modes of posterior densities and called HPD (highest posterior density) estimators. They use different priors to obtain these estimators and it observed that the estimator obtained by using student-t distribution as prior are superior then others as they automatically adjust to the sample dispersion and ignore inconsistent information. He also discusses the point that if the posterior density is bimodal then there is a clear conflict between sample and prior information. At the end results are verified by using the simulation approach. It is concluded the heavy-tailed distributions for sample or priors that are automatically adjust outliers will provide the better inferences then that obtained by using conjugate priors.

Bermudez et. al., (2008) describe the behavior of consumers when they faced with two choices. Since in classical logit model we study the feature of symmetric link but this do not provide good fits for data when one response is much more frequent then the other; so in this paper they use an asymmetric or skewed logit link, proposed by Chen et. al., (1999) to fit a fraud data base from the Spanish insurance market. They use Gibbs sampling and data augmentation for Bayesian analysis of this model. It is observed in results that the use of a skewed link notably improves the percentage of cases that are correctly classified after the model estimation.

Liesenfeld & Richard (2009) propose a generic procedure known as efficient importance sampling (EIC) for the evaluation of likelihood functions for the probit models with correlated errors. Their EIS algorithm covers the standard GHK (Geweke (1991), Hajivassilious (1990), Keane (1994) simulation technique) probability simulator as a special case. They also perform a set of Monte-Carlo experiments in order to illustrate the relative performance of both procedures for the estimation of multinomial multi period probit models. They provide results that are indicating substantial numerical efficiency gain of ML estimators based on GHK-EIS relative to those obtained by using GHK. The evaluation of discrete choice probit models with correlated error terms was first introduced by Thurstone (1927) and applied by Hausman and Wise (1978) to transit choice problems. They use ML integration proposed by Geweke and Keane (2001) to study likelihood function of probit model with correlated error terms that are frequently high-dimensional truncated integral of multivariate normal distribution. They concluded that GHK-EIS provide a significant numerical efficiency gain in ML estimator as compared to GHK.

Rijmen (2008) proposed logistic regression techniques that can be use to restrict the conditional probabilities of a Bayesian network for discrete variables, when all the main effects and interaction between the parent variables are incorporated as covariates. The conditional probabilities are estimated without restrictions as it is a traditional Bayesian network. They also use the ordered logistic regression with ordered categories of the variables, which resulted in more parsimonious model. Then the posterior parameters are estimated by using the modified junction tree algorithm. The main focus of this paper is to learn the parameters of an inferred Bayesian networks for discrete variables, where dependence of relations are encoded through direct edges, more specifically they show how the number of effective parameters of the network can be reduced by adopting a logistic regression frame work for modeling the conditional dependence relations.

Chapter 4

Bayesian Inference of Binary Logistic Regression Model without Intercept

4.1 Introduction

In this chapter, we present the Bayesian analysis of logistic regression model without intercept under informative and noninformative priors. Section 2 gives the introduction to logistic regression model with its different forms. Sections 3, 4 & 5 deal with the derivation of different priors that are used in our study. Section 6 provides the data set and explanation of variables that is used in our research given in Table 4.1. The derivation of posterior distributions using informative and noninformative priors is given in section 7; these posteriors are for the logistic regression model without intercept. For informative prior we set a range of hyperparameters and select the hyperparameters that have minimum standard error, which are given in section 8. Section 9 consists the Bayesian analysis with informative and noninformative priors, which includes the graphs of parameters, the estimated values of parameters and testing for the significance of parameters. Section 10 comprises the classical analysis of logistic regression model without intercept and also the hypothesis testing for the significance of regression coefficient. Section 11 presents the comparison of classical and Bayesian results and their interpretation with respect to the data set given.

4.2 The Binary Logistic Regression Model

Suppose that we have ' n' binomial observations of the form y_i = 1, 2, 3... n where $E(y_i) = p_i$ and p_i is the success probability corresponding to the ith observation. The linear Logistic model for the dependence p_i on the values of the kth explanatory variables x_{1i}, x_{2i}, ..., x_{ni} associated with that observation without intercept is,

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} \tag{4.1}$$

If we have only one explanatory variable for the above logit model then the model become:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_1 x_{1i}$$

If we take exponentiation on both side of the equation (4.1), we get:

$$\exp\left(\log\left(\frac{p_i}{1-p_i}\right)\right) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})$$

Sine on the L.H.S the log will be vanishing with exponential then we have:

$$\frac{p_i}{1-p_i} = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})$$

$$p_i = \frac{\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})}{1+\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})}\ .$$

$$p_i = \frac{1}{1+\exp(-\beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki})} \tag{4.2}$$

If we assume that $\theta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$

Then,

$$p_i = \frac{1}{1+\exp(-\theta_i)} \tag{4.3}$$

Since p_i is the probability of success corresponding to the i[th] observation, while $x_{1i}, x_{2i}, x_{3i}, \ldots, x_{ki}$ are the explanatory Variables with $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ their respective slope coefficients. The shape of given model indicates that the value obtained after estimating the coefficients and for a particular value of explanatory variable it will remain within 0 and 1 that meet the definition of probability theory.

4.3 Uniform Prior

In this section we discuss about the noninformative prior for Bayesian analysis. For more general problems, various suggestions have been advanced for determining a noninformative prior. Noninformative prior, by which we mean a prior that, contains no information about a parameter. For example, when tossing a coin, the probability of ½ to each outcome is clearly noninformative. The simplest situation is to assign each element uniform probability. This is routinely done by Laplace (1812). The uniform prior for the parameter β is given as:

$$p(\beta) \propto 1 \qquad -\infty < \beta < \infty \tag{4.4}$$

4.4 Haldane Prior

Some attempts have been made at finding a priori probabilities, i.e. probability distributions in some sense logically required by the nature of one's state of uncertainty; these are a subject of philosophical controversy. For example (Jaynes 1968) has published an argument based on Lie groups that suggests that the prior for the proportion p of voters voting for a candidate, given no other information, should be the Haldane prior $p^{-1}(1-p)^{-1}$. Haldane prior is the improper when all the parameters are 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 variable have only two categories as yes or no then it will be a Bernoulli trial. So, the Haldane prior will be as:

$$p_H(\beta) \propto \det\{I(\beta)\} \tag{4.5}$$

Where 'det' denotes the determinant and $I(\beta)$ denotes the $n \times n$ Fisher information matrix which is the logarithm of maximum likelihood function of parameter β and partially differentiating twice with respect to the parameter is given below:

$$I(\beta) = -E\left\{\frac{\partial^2 \ln L(\beta)}{\partial \beta^2}\right\} \tag{4.6}$$

Where E stand for the expectation of data:

The Haldane prior has been criticized on the grounds that it yields a posterior distribution that puts 100% of the probability content at either $p = 0$ or at $p = 1$ if a finite sample of voters all favor the same candidate, even though mathematically the posterior probability is simply not defined and thus we cannot even speak of a probability content. The Jeffreys prior $p^{-1/2}(1-p)^{-1/2}$ is therefore preferred.

4.5 Jeffreys prior

Jeffreys (1946, 1961) proposes a noninformative prior. Berger (1985) argues that Bayesian analysis using noninformative prior is the single most powerful method of statistical analysis. The main feature of Jeffreys prior is that it is a uniform measure in information metric, which can be regarded as the natural metric for statistical inference.

Jeffreys rule is defined as the density of the parameters proportional to the square root of the determinant of the Fisher information matrix, symbolically, let

$\beta = (\beta_1, \beta_2, ..., \beta_n)'$ is a vector of parameters $\beta_1, \beta_2, ..., \beta_n$. The prior distribution from the Jeffreys rule is known as Jeffreys prior which is obtained as:

$$p_J(\beta) \propto \sqrt{\det\{I(\beta)\}} \tag{4.7}$$

Where 'det' denotes the determinant and $I(\beta)$ denotes the $n \times n$ Fisher information matrix which is the logarithm of maximum likelihood function of parameter β and partially differentiating twice with respect to the parameter β is given below:

$$I(\beta) = -E\left\{\frac{\partial^2 \ln L(\beta)}{\partial \beta^2}\right\}$$

Where E stands for expectation on data:

4.6 Data Set used in Bayesian Logistic Regression Analysis

The data set for Bayesian analysis of Logistic Regression 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. 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 ≥ 20). Here in this chapter we consider fibrinogen a single explanatory variable and check its individual effect on dependent variable (ESR).

y_i = The Erythrocyte Sedimentation Rate (ESR)

x_{fi} = The amount of protein plasma fibrinogen

Table 4.1: Data

Serial No.	ESR (mm/h) y_i	Fibrinogen (gm/l) x_{fi}	Serial No..	ESR (mm/h) y_i	Fibrinogen (gm/l) x_{fi}
1	0	2.52	17	1	3.53
2	0	2.56	18	0	2.68
3	0	2.19	19	0	2.60
4	0	2.18	20	0	2.23
5	0	3.41	21	0	2.88
6	0	2.46	22	0	2.65
7	0	3.22	23	1	2.09
8	0	2.21	24	0	2.28
9	0	3.15	25	0	2.67
10	0	2.60	26	0	2.29
11	0	2.29	27	0	2.15
12	0	2.35	28	0	2.54
13	1	5.06	29	1	3.93
14	1	3.34	30	0	3.34
15	1	2.38	31	0	2.99
16	1	3.15	32	0	3.32

4.7 Posterior Distribution for the Parameter of the Logistic Regression Without Intercept

Here we consider the simple case of Logistic Regression Model without intercept as:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta x_{fi} \tag{4.8}$$

Here p_i is the probability of success for response variable for i^{th} observation. Where the response variable y_i follows Bernoulli distribution:

So we can also represent the Logit model as.

$$p_r(y_i = 1) = p_i = \frac{1}{1+\exp(-\beta x_{fi})} \tag{4.9}$$

Then the Posterior distribution of the parameter β is defined as:

$$p(\beta \mid data) \propto l(\beta \mid data) \times p(\beta) \tag{4.10}$$

Where $l(\beta \mid data) = \log L(\beta \mid data)$

Now we need to determine the likelihood function and decide upon $p(\beta)$ for the above model.

4.7.1 The Likelihood Function

The likelihood of the i^{th} observation is its probability density function as a function of the parameter β, where (y_i, x_{fi}) are fixed at the observed values. The observations are all independent, now for the given case we precede as follows.

Let y_i be the response variable that is binary in nature i.e. it takes only two values 0 and 1 for 'n' observations. Since the analysis of binary response variable in classical approach the Maximum Likelihood Method (MLE) is used to estimate the unknown parameters of the Binary Logistic Regression Model. However the estimates based on the classical approach are not accurate when the sample size is small. In this situation Bayesian approach provide better and most accurate results. Then if y_i is the response variable and x_{fi}'s the explanatory variables that can be either qualitative or quantitative in nature while p_i is the probability of success corresponding to the i^{th} observation then the probability function is as follows.

$$f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i} \tag{4.11}$$

So the likelihood of whole sample of all observations is the product of the likelihood:

$$L(\beta \mid data) = \prod_{i=1}^{n}\left\{p_i^{y_i}(1-p_i)^{1-y_i}\right\} \tag{4.12}$$

We know that while modeling the binary data, the outcome y_i has a Bernoulli distribution with probability of success p_i that depend upon a set of explanatory variables for a specific i^{th} observation. The probability p_i is regressed on the covariates through a link function that preserves the properties of probability. So, $p_i = H(\beta x_i)$ where x_i the vector of covariates is associated with the i^th^ observation, since $0 \le H(\cdot) \le 1$, and $H(\cdot)$ is a continuous non-decreasing function (Groenewald 2005), it can also be seen in Cox (1971) or Maddala (1983). So the above likelihood function could be written as:

$$L(\beta \mid data) = \prod_{i=1}^{n} \left\{ H(\beta x_i)^{y_i} (1 - H(\beta x_i))^{1-y_i} \right\} \tag{4.13}$$

Now taking log on both sides of equation (4.16) we get the Log likelihood function as:

$$l(\beta \mid data) = \sum_{i=1}^{n} \{ y_i \log H(\beta x_i) + (1 - y_i) \log(1 - H(\beta x_i)) \}$$

Since we know that $p_i = H(\beta x_i)$ then the above log likelihood function becomes:

$$l(\beta \mid data) = \sum_{i=1}^{n} \{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \} \tag{4.14}$$

While p_i is the probability of success for i^{th} observation in data set. Here we use the Log-likelihood function instead to simple likelihood function because the simple likelihood function is too difficult to handle for further manipulation so to make the further manipulation simple we use the log likelihood function. This is routinely used by the Bayesian Econometrician i.e. Chen & Ibrahim (2003), Cengiz et. al., (2001), King & Ryan (2002), Denis Conniffe (1997), Crowder & Sweeting (1989) etc. used log-likelihood function for Bayesian inferences.. So as for as the posterior modes are concern we will use the log likelihood function to construct a posterior distribution for different informative and noninformative priors:

4.7.2 The Prior Distribution

We consider Noninformative and Informative Priors of β in the following sections:

4.7.2.1 The Informative (Normal) Prior

Now we consider the Informative Prior of β as Normal distribution having parameters $mean = a, \text{var}\,iance = b$, so

$$p(\beta) \propto \exp\left\{-\frac{1}{2b}(\beta - a)^2\right\} \quad -\infty < \beta < \infty,\ -\infty < a < \infty \quad b > 0 \tag{4.15}$$

4.7.2.2 The Noninformative (Haldane) prior

The noninformative Prior of β using Haldane Prior can be derived as:

Let us consider the log likelihood function given in equation (4.14).

$$l(\beta \mid data) = \sum_{i=1}^{n}\{y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\}$$

Differentiating the above given form with respect to p_i

$$\frac{\partial l(\beta \mid data)}{\partial p_i} = \frac{\partial}{\partial p_i}\sum_{i=1}^{n}\{y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\}$$

$$= \sum_{i=1}^{n}\left\{y_i\left(\frac{1}{p_i}\right) + (1 - y_i)\left(\frac{1}{1-p_i}\right)(-1)\right\}$$

Again differentiating with respect to p_i:

$$\frac{\partial^2 l(\beta \mid data)}{\partial p_i^2} = \frac{\partial}{\partial p_i^2}\sum_{i=1}^{n}\left\{\left(\frac{y_i}{p_i}\right) - \left(\frac{1-y_i}{1-p_i}\right)\right\}$$

$$= \sum_{i=1}^{n}\left\{\left(\frac{-y_i}{p_i^2}\right) - \left(\frac{1-y_i}{(1-p_i)^2}\right)\right\}$$

As

$$I(\beta) = -E\left\{\frac{\partial^2 l(\beta; y_i)}{\partial p_i^{\ 2}}\right\}$$

Then

$$-E\left[\frac{\partial^2 l(\beta \mid data)}{\partial p_i^{\ 2}}\right] = -E\left[\sum_{i=1}^{n}\left\{\left(\frac{-y_i}{p_i^2}\right) - \left(\frac{1-y_i}{(1-p_i)^2}\right)\right\}\right] \tag{4.16}$$

$$I(\beta) = \left\{\frac{1}{p_i} + \frac{1}{1-p_i}\right\}$$

As $p(\beta) \propto I(\beta)$ then

$$p(\beta) \propto p_i^{\ -1}(1-p_i)^{-1} \tag{4.17}$$

Where

$$p_i = \frac{1}{1+\exp(-\beta x_{fi})}$$

The above equation (4.17) can be written as:

$$p(\beta) \propto H(\beta x_{fi})^{-1}(1-H(\beta x_{fi}))^{-1} \tag{4.18}$$

4.7.2.3 The Noninformative (Jeffreys) Prior

The Noninformative Prior using Jeffreys Prior can be derived as:

Let us again consider the Log likelihood function given in equation (4.14):

$$l(\beta \mid data) = \sum_{i=1}^{n}\{y_i \log(p_i) + (1-y_i)\log(1-p_i)\}$$

Since we know that

$$p_J(\beta) \propto \sqrt{det\{I(\beta)\}}$$

Then

$$p(\beta) \propto p_i^{\ -\frac{1}{2}}(1-p_i)^{-\frac{1}{2}} \tag{4.19}$$

Where

$$p_i = \frac{1}{1+\exp(-\beta x_{fi})}$$

The above equation (4.19) can also be written as:

$$p(\beta) \propto H(\beta x_{fi})^{-\frac{1}{2}}(1-H(\beta x_{fi}))^{-\frac{1}{2}} \qquad (4.20)$$

4.7.2.4 The Noninformative (Uniform) Prior

We consider noninformative prior of β as Uniform Prior as; see section (4.4) equation (4.4):

$$p(\beta) \propto 1 \qquad \beta \in (-\infty, \infty) \qquad (4.21)$$

4.7.3 The Posterior Distribution

The Posterior distributions of β using noninformative and informative priors are given in the following sections:

4.7.3.1 The Posterior Distribution Using Normal Prior

The posterior distribution of β using the Normal prior distribution, for this we consider Log likelihood function (4.14) and Normal Prior (4.15):

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} \exp\left\{-\frac{1}{2b}(\beta-a)^2\right\}$$

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \left\{y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right\} \exp\left\{-\frac{1}{2b}(\beta-a)^2\right\} \qquad (4.22)$$

Since we know that the simple Logistic Regression Model without intercept and having only one explanatory variable is as follows; see section (4.7):

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta x_{fi}$$

While $p_i = \frac{1}{1+\exp(-\beta x_{fi})}$

Or $$1-p_i = \frac{1}{1+\exp(\beta x_{fi})} \tag{4.23}$$

Taking log on both sides of (4.23) we get:

$$\log(1-p_i) = -\log(1+\exp(\beta x_{fi})) \tag{4.24}$$

Now replace equation (4.8) and (4.24) in equation (4.22) we get:

$$p(\beta \mid data) \propto \sum_{i=1}^{n}\{y_i\beta x_{fi} - \log(1+\exp(\beta x_{fi}))\}\exp\left\{-\frac{1}{2b}(\beta-a)^2\right\}$$

$$p(\beta \mid data) = \frac{1}{k}\sum_{i=1}^{n}\{y_i\beta x_{fi} - \log(1+\exp(\beta x_{fi}))\}\exp\left\{-\frac{1}{2b}(\beta-a)^2\right\}$$

Let us suppose that $\phi = (\beta - a)^2$

$$p(\beta \mid data) = \frac{1}{k}\sum_{i=1}^{n}\{y_i\beta x_{fi} - \log(1+\exp(\beta x_{fi}))\}\exp\left\{-\frac{\phi}{2b}\right\} \tag{4.25}$$

$$-\infty < \beta < \infty$$

This is the posterior distribution of β, where k is the normalizing constant. Here our main objective is to estimate β. Then for this purpose if we partially differentiate the above equation (4.25) with respect to β and equate it to zero. So this numerical solution will provide us the Posterior Mode, so for this we proceed as follows.

Differentiate (4.25) with respect to β we get:

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \frac{\partial}{\partial \beta}\sum_{i=1}^{n}\{y_i\beta x_{fi} - \log(1+\exp(\beta x_{fi}))\}\exp\left\{-\frac{1}{2b}(\beta-a)^2\right\}$$

$$= \sum_{i=1}^{n} \left\{ \exp\left\{ -\frac{\phi}{2b} \right\} \left\{ y_i x_{fi} - \frac{x_{fi}\exp(-\beta x_{fi})}{(1+\exp(-\beta x_{fi}))^2 \left(1 - \frac{1}{1+\exp(-\beta x_{fi})}\right)} \right\} - \right.$$

$$\left. \left\{ y_i \beta x_{fi} + \log\left(1 - \frac{1}{1+\exp(-\beta x_{fi})}\right) \right\} \exp\left\{ -\frac{\phi}{2b} \right\} \frac{1}{b}(\beta - a) \right\}$$

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \sum_{i=1}^{n} \left\{ \exp\left\{ -\frac{\phi}{2b} \right\} \left\{ \left(y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right) - \right. \right.$$

$$\left. \left. \frac{1}{b}(\beta - a) \left\{ y_i \beta x_{fi} + \log\left(1 - \frac{1}{1+\exp(-\beta x_{fi})}\right) \right\} \right\} \right\}$$

Now put $\frac{\partial p(\beta \mid data)}{\partial \beta} = 0$ while $\lambda = \frac{1}{b}(\beta - a)$

$$\sum_{i=1}^{n} \left\{ \exp\left\{ -\frac{\phi}{2b} \right\} \left\{ \left(y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right) - \lambda \left\{ y_i \beta x_{fi} + \log\left(\frac{1}{1+\exp(\beta x_{fi})} \right) \right\} \right\} \right\} = 0 \tag{4.26}$$

By solving this numerically, the posterior mode of β can be obtained.

4.7.3.2 The Posterior Distribution Using Haldane Prior

The posterior distribution of β using the Haldane prior distribution, we consider Log likelihood function (4.14) and Haldane Prior (4.17) then

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \{ y_i \log(p_i) + (1 - y_i)\log(1 - p_i) \} p_i^{-1}(1 - p_i)^{-1}$$

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{p_i}{1 - p_i} \right) + \log(1 - p_i) \right\} p_i^{-1}(1 - p_i)^{-1} \tag{4.27}$$

As we know that $p_i = \frac{1}{1+\exp(-\beta x_{fi})}$

Then $p_i^{-1} = 1 + \exp(-\beta x_{fi})$ and $(1 - p_i)^{-1} = 1 + \exp(\beta x_{fi})$ (4.28)

Now after replacing (4.8), (4.24) and (4.28) in equation (4.27) we obtain:

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \{y_i \beta x_{fi} - \log(1 + \exp(\beta x_{fi}))\}(1 + \exp(-\beta x_{fi}))(1 + \exp(\beta x_{fi}))$$

$$p(\beta \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\{y_i \beta x_{fi} - \log(1 + \exp(\beta x_{fi}))\} \{2 + \exp(-\beta x_{fi}) + \exp(\beta x_{fi})\} \right] \quad (4.29)$$

$$-\infty < \beta < \infty$$

This is the posterior distribution of β, where k is the normalizing constant. We will use the above posterior distribution to obtain the estimated value of parameter β. Then for this purpose if we partially differentiate the above equation (4.29) with respect to β and equate it to zero. So this numerical solution will provide us the Posterior Mode, so for this we proceed as follows.

Differentiate (4.29) with respect to β we get:

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \left[\{y_i \beta x_{fi} - \log(1 + \exp(\beta x_{fi}))\} \{2 + \exp(-\beta x_{fi}) + \exp(\beta x_{fi})\} \right]$$

$$= \sum_{i=1}^{n} \left\{ (1 + \exp(-\beta x_{fi}))(1 + \exp(\beta x_{fi})) \left\{ y_i x_{fi} - \frac{x_{fi} \exp(-\beta x_{fi})}{(1 + \exp(-\beta x_{fi}))^2 \left(1 - \frac{1}{1 + \exp(-\beta x_{fi})} \right)} \right\} + \right.$$

$$\exp(\beta x_{fi})(1 + \exp(-\beta x_{fi})) x_{fi} \left\{ y_i \beta x_{fi} - \log\left(1 - \frac{1}{1 + \exp(-\beta x_{fi})} \right) \right\} -$$

$$\left. (1 + \exp(\beta x_{fi})) \exp(-\beta x_{fi}) x_{fi} \left\{ y_i \beta x_{fi} - \log\left(1 - \frac{1}{1 + \exp(-\beta x_{fi})} \right) \right\} \right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\beta x_{fi}))(1+\exp(\beta x_{fi}))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$\exp(\beta x_{fi})(1+\exp(-\beta x_{fi}))x_{fi}\left\{y_i\beta x_{fi}-\log\left(\frac{\exp(-\beta x_{fi})}{1+\exp(-\beta x_{fi})}\right)\right\}-$$

$$\left.(1+\exp(\beta x_{fi}))\exp(-\beta x_{fi})x_{fi}\left\{y_i\beta x_{fi}-\log\left(\frac{\exp(-\beta x_{fi})}{1+\exp(-\beta x_{fi})}\right)\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\beta x_{fi}))(1+\exp(\beta x_{fi}))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(\beta x_{fi})}\right\}+\right.$$

$$\exp(\beta x_{fi})(1+\exp(-\beta x_{fi}))x_{fi}\left\{y_i\beta x_{fi}+\log(1+\exp(\beta x_{fi}))\right\}-$$

$$\left.\exp(-\beta x_{fi})(1+\exp(\beta x_{fi}))x_{fi}\left\{y_i\beta x_{fi}+\log(1+\exp(\beta x_{fi}))\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\beta x_{fi}))(1+\exp(\beta x_{fi}))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$\left\{\exp(\beta x_{fi})(1+\exp(-\beta x_{fi}))-\exp(-\beta x_{fi})(1+\exp(\beta x_{fi}))\right\}$$

$$\left.x_{fi}\left\{y_i\beta x_{fi}+\log(1+\exp(\beta x_{fi}))\right\}\right\}$$

$$\frac{\partial p(\beta\mid data)}{\partial\beta}=\sum_{i=1}^{n}\left\{(2+\exp(-\beta x_{fi})+\exp(\beta x_{fi}))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$\left.x_{fi}\left\{y_i\beta x_{fi}+\log(1+\exp(\beta x_{fi}))\right\}(\exp(\beta x_{fi})-\exp(-\beta x_{fi}))\right\}$$

Now put $\dfrac{\partial p(\beta\mid data)}{\partial\beta}=0$

$$\sum_{i=1}^{n}\left\{(2+\exp(-\beta x_{fi})+\exp(\beta x_{fi}))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$x_{fi}\left\{y_i\beta x_{fi}+\log\left(1+\exp\left(\beta x_{fi}\right)\right)\right\}\left(\exp\left(\beta x_{fi}\right)-\exp\left(-\beta x_{fi}\right)\right)\Big\}=0 \quad (4.30)$$

Solving numerically the above equation, the posterior mode of β can be obtained.

4.7.3.3 The Posterior Distribution Using Jeffreys Prior

The posterior distribution of β using the Jeffreys prior distribution, we consider Log likelihood function (4.14) and the Jeffreys Prior (4.19):

$$p(\beta\,|\,data)\propto\sum_{i=1}^{n}\{y_i\log(p_i)+(1-y_i)\log(1-p_i)\}p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}}$$

$$p(\beta\,|\,data)\propto\sum_{i=1}^{n}\left\{y_i\log\left(\frac{p_i}{1-p_i}\right)+\log(1-p_i)\right\}p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}} \quad (4.31)$$

As we know that $p_i=\dfrac{1}{1+\exp(-\beta x_{fi})}$

Then $p_i^{-\frac{1}{2}}=\sqrt{1+\exp(-\beta x_{fi})}$ and $(1-p_i)^{-\frac{1}{2}}=\sqrt{1+\exp(\beta x_{fi})}$ (4.32)

Now after replacing equations (4.8), (4.24) and (4.32) in (4.31) we get:

$$p(\beta\,|\,data)\propto\sum_{i=1}^{n}\{y_i\beta x_{fi}-\log(1+\exp(\beta x_{fi}))\}\sqrt{1+\exp(-\beta x_{fi})}\sqrt{1+\exp(\beta x_{fi})}$$

$$p(\beta\,|\,data)=\frac{1}{k}\sum_{i=1}^{n}\left[\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})}\left\{y_i\beta x_{fi}-\log(1+\exp(\beta x_{fi}))\right\}\right] \quad (4.33)$$

$$-\infty<\beta<\infty$$

This is the posterior distribution of β, where k is the normalizing constant. Here our main objective is to estimate β. For this purpose, we partially differentiate the above equation (4.33) with respect to β and equate it to zero. So this numerical solution will provide us the posterior mode, so for this we proceed as follows:

Differentiate (4.33) with respect to β we get:

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \frac{\partial}{\partial \beta}\sum_{i=1}^{n}\left[\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})}\left\{y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi}))\right\}\right]$$

$$= \sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\beta x_{fi})}\sqrt{1+\exp(\beta x_{fi})}\left\{y_i x_{fi} - \frac{x_{fi}\exp\left(-\beta x_{fi}\right)}{(1+\exp(\beta x_{fi}))^2\left(1-\frac{1}{1+\exp(-\beta x_{fi})}\right)}\right\}+\right.$$

$$\frac{x_{fi}\exp(\beta x_{fi})\sqrt{1+\exp(-\beta x_{fi})}\left\{y_i \beta x_{fi} + \log\left(1-\frac{1}{1+\exp(-\beta x_{fi})}\right)\right\}}{2\sqrt{1+\exp(\beta x_{fi})}} -$$

$$\left.\frac{x_{fi}\exp(-\beta x_{fi})\sqrt{1+\exp(\beta x_{fi})}\left\{y_i \beta x_{fi} + \log\left(1-\frac{1}{1+\exp(-\beta x_{fi})}\right)\right\}}{2\sqrt{1+\exp(-\beta x_{fi})}}\right\}$$

$$= \sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\beta x_{fi})}\sqrt{1+\exp(\beta x_{fi})}\left\{y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$\frac{x_{fi}\exp(\beta x_{fi})\sqrt{1+\exp(-\beta x_{fi})}\left\{y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi}))\right\}}{2\sqrt{1+\exp(\beta x_{fi})}} -$$

$$\left.\frac{x_{fi}\exp(-\beta x_{fi})\sqrt{1+\exp(\beta x_{fi})}\left\{y_i \beta x_{fi} - \log\log(1+\exp(\beta x_{fi}))\right\}}{2\sqrt{1+\exp(-\beta x_{fi})}}\right\}$$

$$= \sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})}\left\{y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})}\right\}+\right.$$

$$\left\{\frac{\exp(\beta x_{fi})\sqrt{1+\exp(-\beta x_{fi})}}{2\sqrt{1+\exp(\beta x_{fi})}} - \frac{\exp(-\beta x_{fi})\sqrt{1+\exp(\beta x_{fi})}}{2\sqrt{1+\exp(-\beta x_{fi})}}\right\}$$

$$\left. x_{fi}\left\{y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi}))\right\}\right\}$$

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \sum_{i=1}^{n} \left\{ \sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right\} + \right.$$

$$\left. x_{fi} \left\{ y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi})) \right\} \left\{ \frac{\exp(\beta x_{fi}) - \exp(-\beta x_{fi})}{2\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})}} \right\} \right\}$$

Now put $\frac{\partial p(\beta \mid data)}{\partial \beta} = 0$

$$\sum_{i=1}^{n} \left\{ \sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right\} + \right.$$

$$\left. x_{fi} \left\{ y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi})) \right\} \left\{ \frac{\exp(\beta x_{fi}) - \exp(-\beta x_{fi})}{2\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})}} \right\} \right\} = 0 \qquad (4.34)$$

The posterior mode can be obtained by solving the above equation (4.34) numerically.

4.7.3.4 The Posterior distribution Using Uniform Prior

Now using the Log likelihood function (4.14) and the Uniform prior distribution (4.21), the Posterior distribution of β is found to be:

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \{ y_i \log(p_i) + (1-y_i)\log(1-p_i) \} 1$$

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{p_i}{1-p_i} \right) + \log(1-p_i) \right\} \qquad (4.35)$$

Replacing equations (4.8) and (4.24) in (4.35) we get:

$$p(\beta \mid data) \propto \sum_{i=1}^{n} \{ y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi})) \}$$

$$p(\beta \mid data) = \frac{1}{k} \sum_{i=1}^{n} \{ y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi})) \} \qquad -\infty < \beta < \infty \qquad (4.36)$$

This is the posterior distribution of β, with normalizing constant k. We proceed further to obtain the estimate of parameter β. Then partially differentiate the above equation (4.36) with respect to β and equate it to zero. That can be further solved numerically to obtain the posterior mode, so for this we proceed as follows.

Differentiate (4.36) with respect to β we get:

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \{ y_i \beta x_{fi} - \log(1+\exp(-\beta x_{fi})) \}$$

$$= \sum_{i=1}^{n} \left\{ y_i x_{fi} - \left(\frac{1}{1+\exp(\beta x_{fi})} \cdot \frac{\partial}{\partial \beta} (1+\exp(\beta x_{fi})) \right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i x_{fi} - \left(\frac{\exp(\beta x_{fi})}{1+\exp(\beta x_{fi})} \right) x_{fi} \right\}$$

$$\frac{\partial p(\beta \mid data)}{\partial \beta} = \sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right\}$$

Now $\frac{\partial p(\beta \mid data)}{\partial \beta} = 0$

$$\sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta x_{fi})} \right\} = 0 \qquad (4.37)$$

The numerical solution of above equation provides us the posterior mode.

4.8 Selection of Hyperparameters

Since we know that the prior distribution of parameter β is Normal (a,b) and our main objective here is to find the values of these hyperparameters while 'a' is the mean of prior distribution and 'b' is the variance of the prior distribution. The idea of selecting hyperparameters is taken from Bian (1997), they assume Normal & Student-t priors with mean zero and decide about the posterior distribution at different values of variance for logit model, they also take the logit model without intercept and with intercept. But we have suggest rang of values for both parameters and select the values with minimum standard error. Since the next observation only depend on the parameter β through the function $\log\left(\frac{p_i}{1-p_i}\right) = \beta x_{fi}$. We have suggested a range of values of hyperparameters by observing the variation in regression coefficient and also the variable of interest. The values are given as follows:

Table 4.2

Posterior Estimates at Different Values of Hyperparameters

Hyperparameters		Posterior Mode	Posterior Mean	Standard Error
Mean a_1	*Variance* b_1	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$
0	1	−0.356785	−0.40421	0.149078
8.50	2.50	−0.724716	−4.52556	2.50413
7.50	2	−0.698402	−4.46906	2.45921
6.50	1.50	−0.668756	−4.43835	2.30297
5.50	1	−0.664757	−4.17370	2.14357
4.50	0.90	−0.594808	−3.03357	1.27286
3.50	0.80	−0.546194	−2.02769	0.95781
2.50	0.70	−0.483658	−1.18625	0.87247
2	0.60	−0.443846	−0.85877	0.75325
1.90	0.50	−0.434886	−0.83030	0.70124
1.80	0.40	−0.425520	−0.79985	0.65941
1.70	0.30	−0.415706	−0.76746	0.57294
1.60	0.20	−0.405397	−0.73318	0.50451
1.50	0.20	−0.394535	−0.66310	0.43129
1.40	0.20	−0.383055	−0.59655	0.35547
1.30	0.20	−0.370875	−0.53354	0.29579
1.20	0.20	−0.357898	−0.47407	0.23480
1.10	0.20	−0.344005	−0.41814	0.18542
1	0.20	−0.325645	−0.34024	0.13872

Where $mean = a$ and $variance = b$ for the prior distribution. We suggest different values for the hyperparameters and find the values of posterior estimates. So finally we decided to select the values of hyperparameters as $mean = 1$ and $variance = 0.20$ and used these values for further Bayesian analysis because this set (1, 0.20) has the smallest standard error.

4.9 Bayesian Analysis with Informative and Noninformative Priors

In this section we will present the Bayesian analysis with informative and noninformative priors. The analysis is based on the posterior distributions that are derived in previous sections:

4.9.1 Bayesian Analysis Using Normal Prior

In this section we will present the Bayesian analysis of logistic regression model without intercept by using informative prior as Normal. Then the Posterior distribution for the parameter β derived in section (4.7.3.1) see equation (4.25):

$$p(\beta \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left\{y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi}))\right\} \exp\left\{-\frac{\phi}{2b}\right\} \quad -\infty < \beta < \infty$$

where k is the normalizing constant:

The graph of the posterior density of the parameter β is shown using the data set given in Table 4.1.

The Graph of Posterior Density Using Informative Prior

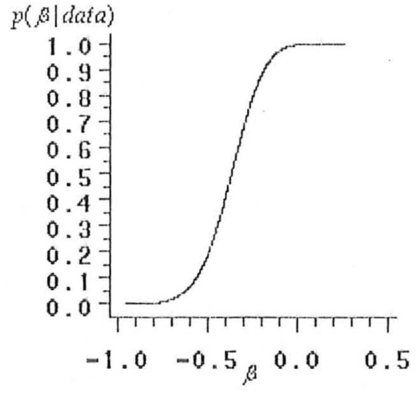

Figure 1

The graph of the posterior density of β under informative prior shows that it is the Cumulative Density Function of logistic distribution:

4.9.1.1 Posterior Estimates

We have designed programs in SAS package; program is given in appendix I and also a similar program given in appendix IV to obtain the value of standard error of the parameter, while using the data set given in Table 4.1 to obtain posterior mode and the posterior mean and standard error which are given in Table 4.3:

Table 4.3

Posterior Estimates Using Informative Normal Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}$	-0.3402	-0.3256	0.1387	0.7221	-0.1053

Here we observe that the posterior mean is less then the posterior mode which indicates that the distribution of β is not symmetrical as the graph also indicates see figure 1. The Karl Pearson coefficient of skewness (SK_p) is also computed which indicates the level of asymmetry of the posterior distribution of β. We can see in Table 4.3 that the odds ratio is less then 1 which indicates that the variable fibrinogen is less likely to occur. So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately increases ESR by 0.7221. This is very low for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h. So it can be concluded that here fibrinogen is not playing any significant role to increase the level of ESR in any healthy individual.

4.9.1.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the posterior distribution upon the parameter i.e.

We test the following hypotheses:

$H_0 : \beta \geq 0$ Versus $H_1 : \beta < 0$

The posterior probability for H_0 is:

$$p(\beta \geq 0) = \int_0^{\infty} p(\beta \mid data) d\beta$$

Now the posterior probability using informative prior is:

$$p_0 = \int_0^1 \frac{1}{k} \sum_{i=1}^{n} \{ y_i \beta x_{fi} - \log(1 + \exp(\beta x_{fi})) \} \exp\left\{ -\frac{1}{2} (\beta - 0.20)^2 \right\} d\beta$$

A program is designed SAS package, similar prográm is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_0 = 0.0031186$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.312% chance to accept H_0 and we conclude that for this model fibrinogen is not playing any significant role to increase the level of ESR for healthy individual if the level of this protein plasma is increases.

4.9.2 Bayesian Analysis Using Haldane Prior

In this section we will present the Bayesian analysis of logistic regression model without intercept by using noninformative: Haldane prior. The Posterior distribution for the parameter β derived in section (4.7.3.2) see equation (4.29) is:

$$p(\beta \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\{ y_i \beta x_{fi} - \log(1 + \exp(\beta x_{fi})) \} \{ 2 + \exp(-\beta x_{fi}) + \exp(\beta x_{fi}) \} \right]$$

$$-\infty < \beta < \infty$$

where k is the normalizing constant:

The graph of posterior density of the parameter β is shown in figure 2 using the data set given in Table 4.1.

The Graph of Posterior Density Using Haldane Prior

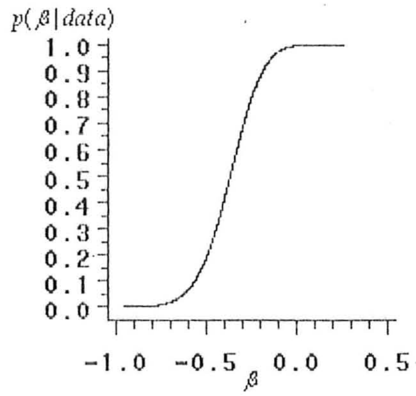

Figure 2

The posterior distribution of β is logistic type as its graph indicates.

4.9.2.1 Posterior Estimates

We have designed programs in SAS package, similar program is given in appendix 1 while using the data set given in Table 4.1 to obtain posterior mode and the posterior mean and standard error which are given in Table 4.3:

Table 4.4

Posterior Estimates Using Noninformative Haldane Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}$	-0.3463	-0.3312	0.1410	0.7181	-0.1071

Here we also observe the same results with a slight difference in the values of posterior estimates. The Karl Pearson coefficient of skewness (SK_p) is also indicates almost same level of asymmetry of the posterior distribution of β. We can see in Table 4.4 that the odds ratio is again less then 1 which indicates that the variable fibrinogen is less likely to occur. This is very low for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h.

4.9.2.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the posterior distribution upon the parameter:

Now the posterior probability using Haldane prior is:

$$p_0 = \int_0^1 \frac{1}{k} \sum_{i=1}^{n} \left[\{ y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi})) \} \{ 2+\exp(-\beta x_{fi}) + \exp(\beta x_{fi}) \} \right] d\beta$$

A program is designed SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.0031571$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.316% chance to accept H_0 and we conclude that for this model fibrinogen is not playing any significant role to increase the level of ESR for healthy individual to become an unhealthy individual if the level of this protein plasma is increases.

4.9.3 Bayesian Analysis Using Jeffreys Prior

In this section we will present the Bayesian analysis of logistic regression model without intercept by using noninformative Jeffreys prior. Then the Posterior distribution for the parameter β derived in section (4.7.3.3) see equation (4.33):

$$p(\beta \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\sqrt{2+\exp(-\beta x_{fi}) + \exp(\beta x_{fi})} \{ y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi})) \} \right]$$

where k is the normalizing constant: $-\infty < \beta < \infty$

The graph of posterior density of the parameter β is shown in figure 3 using the data set given in Table 4.1.

The Graph of Posterior Density using Jeffreys Prior

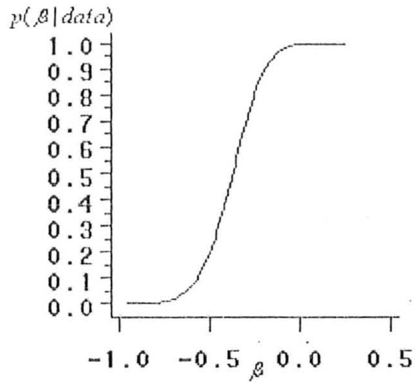

Figure 3

The graph of the posterior distribution of β shows that it is a logistic type and we have used the logistic link. This graph is very much similar to the graphs which are presented in the previous sections.

4.9.3.1 Posterior Estimates

We have designed programs in SAS package; similar program is given in appendix 1 and for standard error similar program is given in appendix IV, while using the data set given in Table 4.1 to obtain posterior mode and the posterior mean and standard error which are given in Table 4.5:

Table 4.5

Posterior Estimates Using Noninformative Jeffreys Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}$	-0.3681	-0.3525	0.1430	0.7029	-0.1091

The results given in Table 5.4 are also same as in previous section with slight difference in the values of posterior estimates. The Karl Pearson coefficient of skewness (SK_p) is also indicates almost same level of asymmetry of the posterior distribution of β. We can see in Table 4.5 that the odds ratio is also less then 1 which gives the same indication as in the previous sections. So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately 0.7029 increases in the level of ESR. This is very low for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h. So it can be concluded that here again fibrinogen is not playing any significant role to increase the level of ESR in any healthy individual.

4.9.3.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the posterior distribution upon the parameter:

Now the posterior probability using Jeffreys prior is:

$$p_0 = \int_0^1 \frac{1}{k} \sum_{i=1}^{n} \left[\sqrt{2+\exp(-\beta x_{fi})+\exp(\beta x_{fi})} \left\{ y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi})) \right\} \right] d\beta$$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.0031749$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.317% chance to accept H_0 and we conclude that for this particular model fibrinogen is not playing any significant role to effect the ESR if this protein is rise in the blood plasma.

4.9.4 Bayesian Analysis Using Uniform Prior

In this section we will describe the Bayesian analysis of logistic regression model without intercept by using noninformative Uniform prior. Then the Posterior distribution for the parameter β derived in section (4.7.3.4) see equation (4.36):

$$p(\beta \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left\{y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi}))\right\} \qquad -\infty < \beta < \infty$$

Where k is the normalizing constant:

The graph of posterior density of the parameter β is shown using the data set in Table 4.1 and design a program in SAS package:

The Graph of Posterior Density using Uniform Prior

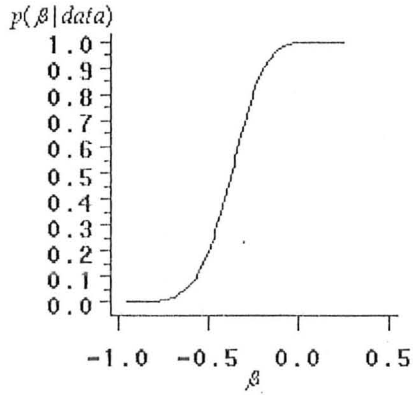

Figure 4

The graph of the posterior distribution of β shows that it is a logistic type as we have used the logistic link. This graph is very much similar to the graphs which are presented in the previous sections.

4.9.4.1 Posterior Estimates

We have designed programs in SAS package, similar program is given in appendix 1 while using the data set given in Table 4.1 to obtain posterior mode and the posterior mean and standard error which are given in Table 4.6:

Table 4.6

Posterior Estimates Using Noninformative Uniform Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}$	-0.3726	-0.3569	0.1434	0.6998	-0.1095

The results given in Table 5.6 are also same as in previous section with slight difference in the values of posterior estimates. The Karl Pearson coefficient of skewness (SK_p) also indicates almost same level of asymmetry of the posterior distribution of β. We can see in Table 4.6 that the odds ratio is also less then 1 which gives the same indication as in the previous sections that the variable (Fibrinogen) is less likely to occur. So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately 0.6998 increases in the level of ESR. This is very low for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h. So it can be concluded that here again fibrinogen is not playing any significant role to increase the level of ESR in any healthy individual.

4.9.4.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the posterior distribution upon the parameter:

Now the posterior probability for H_0 using Uniform prior is:

$$p_0 = \frac{1}{k}\int_0^1 \sum_{i=1}^{n} \{y_i \beta x_{fi} - \log(1+\exp(\beta x_{fi}))\} d\beta$$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.0032583$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.326% chance to accept H_0 and we conclude that for this particular model fibrinogen is not playing any significant role to effect the ESR to increase its level for any healthy individual from 20 mm/h to an unhealthy individual with ESR greater equal than 20 mm/h, if the level of this protein is increases in the blood plasma:

4.10 Classical Regression Analysis

For the comparison purpose now we take the classical estimate and test the hypothesis. For this we have simply run the logistic regression without intercept model. Now the classical estimate and Hypothesis testing is given in following section:

4.10.1 Classical Estimate

Using the data set given in Table 4.1 and having run the logistic regression we obtain:

Table 4.7

Output of Logistic Regression Using Classical Approach

Coefficient	Classical Estimate	Standard Error	Z-Statistic	P-Value	Odds Ratio
$\hat{\beta}$	-0.3569	0.1434	-2.4894	0.0128	0.6998

4.10.2 Classical way of Hypothesis Testing

We have the logistic regression model as:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta x_{fi}$$

Hypothesis

$H_0 \geq 0$ Versus $H_1 < 0$

Since $\beta = -0.3569$ and standard error of β is 0.1434 then the value of Wald t-statistic is, $t = -2.4894$

Since the p-value from this regression is 0.0128, it indicates that we accept H_0 at 1% level of significance and do not accept H_0 at any other level of significance. So it can be concluded that fibrinogen is playing significant role at 5% level of significance and it effect the ESR if this protein (Fibrinogen) increases in the blood plasma:

4.11 Comparison of Bayesian and Classical Logistic Regression Analysis

We compare the results obtained by using Bayesian and Classical techniques. The results are presented in Table 4.8 using different priors these, results can be compared with the results given in table 4.7:

Table 4.8

Posterior Estimates for Without Intercept Logistic Regression Model

Coefficient $\hat{\beta}$	Noninformative Prior			Informative Prior
	Uniform Prior	Jeffreys Prior	Haldane Prior	
Posterior Mode	−0.3569	−0.3525	−0.3312	−0.3256
Posterior Mean	−0.3726	−0.3681	−0.3463	−0.3402
Odds Ratio	0.6998	0.7029	0.7181	0.7221
Standard Error	0.1434	0.1430	0.1410	0.1387
SK_p	−0.1095	−0.1091	−0.1071	−0.1053

The results found by using Classical logistic regression and in Bayesian logistic regression with Uniform prior are looks like same i.e. the coefficients, p-value and odds ratio. Here odds ratio interpreted as the approximated change in the risk of disease for every one unit increase in the amount of fibrinogen. We observe here one thing that with this type of logistic model the results by using Bayesian approach with all prior say nothing about the significance of the parameter so is indicated by interpretation of odds ratio but classical results are giving evidence in favor of the regression coefficient in term of p-value if we select 5% as level of significance although which is too high with these type of experiments, contradict with the value of odds ratio which is not the case in Bayesian. So it can be said that results are much accurate by using Bayesian approach, while the results are much improved with Haldane and informative (Normal) prior as compared to uniform and Jeffreys.

Chapter 5

Bayesian Inference of Logistic Regression Model with Intercept

5.1 Introduction

In this chapter, we deal with the Bayesian analysis of logistic regression model with intercept using one explanatory variable for response binary variable, under informative and noninformative priors. Section 2 consists the introduction to logistic regression model with its different forms. Section 3 presents the joint posterior model and joint likelihood function also the derivation of different joint priors that are used in our study. The derivation of posterior distribution using informative and noninformative prior are also given, we use the log likelihood function to obtain the posterior distribution. For informative prior we set a range of hyperparameters and select the hyperparameters that have minimum standard errors for parameters, these are present in section 4. Section 5 comprises the Bayesian analysis with informative and noninformative priors, which includes the graphs and posterior estimates and hypotheses testing for the significance of parameter. Section 6 discusses the classical analysis of logistic regression model with intercept and also the hypothesis testing for the significance of regression coefficient. Sections 7 contain the comparison of classical and Bayesian results and their interpretation with respect to the data set given.

5.2 Binary Logistic Regression Model (BLR) with Intercept

Here we will consider the binary logistic regression model with intercept that is given as follows:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{fi} \qquad (5.1)$$

Here β_0 is the intercept and β_1 is the slope coefficient for the explanatory variable fibrinogen. The above logistic regression model can also be represented as:

$$p_i = p_r(y_i = 1) = \frac{1}{1+\exp\{-(\beta_0 + \beta_1 x_{fi})\}} \quad (5.2)$$

5.3 Joint Posterior Distribution of Binary Logistic Regression (BLR) Model with Intercept

Then the joint Posterior distribution of the parameter β_0 and β_1 is defined as:

$$p(\beta_0,\beta_1 \mid data) \propto l(\beta_0,\beta_1 \mid data) \times p(\beta_0,\beta_1) \quad (5.3)$$

Here $p(\beta_0,\beta_1 \mid data)$ is the joint posterior distribution while $l(\beta_0,\beta_1 \mid data)$ is the joint log likelihood function and $p(\beta_0,\beta_1)$ is the joint prior distribution for β_0 and β_1. The prior distributions are considered to be independent, as the independent priors are extensively used in literature, this idea can also be seen in Bian (1997), Dreze (1977), Bian (1989), Bedrick et. al. (1996) etc. So we need to decide upon the joint prior distribution and the joint log likelihood function.

5.3.1 Joint Likelihood Function

The joint likelihood of the i^{th} observation is its probability density function as a function of the two parameters β_0 and β_1 where (y_i, x_{fi}) are fixed at the observed values. The observations are all independent, now for the given case we precede as follows.

Let y_i be the response variable that is binary in nature i.e. it takes only two values 0 and 1 for 'n' observations. Since the analysis of binary response variable in classical approach, the Maximum Likelihood Method (MLE) is used to estimate the unknown parameters of the Binary Logistic Regression Model. However the estimates based on the classical approach are not accurate when the sample size is small. In this situation Bayesian approach provides better and accurate results. Then if y_i is the response variable and x_{fi}'s the explanatory variables that can be either qualitative or quantitative in nature while p_i is the probability of success corresponding to the i[th] observation then the joint likelihood function can be presented as:

$$L(\beta_0, \beta_1 \mid data) = \prod_{i=1}^{n} \left\{ p_i^{y_i} (1 - p_i)^{1-y_i} \right\} \qquad (5.4)$$

Now if $p_i = H(\beta' x_i)$ while $\beta' = (\beta_0, \beta_1)$ then the likelihood function becomes as:

$$L(\beta_0, \beta_1 \mid data) = \prod_{i=1}^{n} \left\{ H(\beta' x_i)^{y_i} (1 - H(\beta' x_i))^{1-y_i} \right\} \qquad (5.5)$$

Taking log on both sides of above function we get the Log likelihood function as:

$$\log L(\beta_0, \beta_1 \mid data) = \sum_{i=1}^{n} \{ y_i \log(H(\beta' x_i)) + (1 - y_i) \log(1 - H(\beta' x_i)) \}$$

Sine we know that $p_i = H(\beta' x_i)$ then for further consideration we can write the above log likelihood function as:

$$l(\beta_0, \beta_1 \mid data) = \sum_{i=1}^{n} \{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \} \qquad (5.6)$$

Where p_i is the probability of success for i[th] observation in data set and be represented in the logistic regression model as:

$$p_i = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_{ji})\}}$$

5.3.2 Joint Prior Distribution

We consider the joint noninformative and informative priors of β_0 and β_1 in the following sections:

5.3.2.1 Joint Informative (Normal) Prior

Here we consider the independent normal prior for each parameter. The joint Informative prior of the two parameters is the product of the two individual priors:

$$p(\beta_0, \beta_1) = p(\beta_0)p(\beta_1)$$

Where $\beta_0 \sim N(a_0, b_0)$ and $\beta_1 \sim N(a_1, b_1)$ here a_0 and a_1 are means while b_0 and b_1 are the variances. Therefore

$$p(\beta_0, \beta_1) \propto \exp\left\{-\frac{1}{2b_0}(\beta_0 - a_0)^2\right\}\exp\left\{-\frac{1}{2b_1}(\beta_1 - a_1)^2\right\}$$

$$p(\beta_0, \beta_1) \propto \exp\left\{-\frac{b_1(\beta_0 - a_0)^2 + b_0(\beta_1 - a_1)^2}{2b_0 b_1}\right\} \qquad -\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty \quad (5.7)$$

5.3.2.2 Joint Noninformative (Haldane) prior

The joint noninformative (Haldane) prior by using the log likelihood function given in (5.6) is derived as:

$$p_H(\beta_0, \beta_1) \propto p_i^{-1}(1 - p_i)^{-1} \qquad (5.8)$$

where $p_i = \dfrac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_{ji})\}}$

The above equation (5.7) can also be written as:

$$p(\beta_0, \beta_1) \propto H(\beta' x_i)^{-1}(1 - H(\beta' x_i))^{-1} \qquad (5.9)$$

5.3.2.3 Joint Noninformative (Jeffreys) Prior

The joint noninformative (Jeffreys) prior by using the log likelihood function given in (5.6) is derived as:

Since we know that

$$p_J(\beta_0, \beta_1) \propto \sqrt{\det|I(\beta_0, \beta_1)|}$$

Then $p(\beta_0, \beta_1) \propto p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}}$ (5.10)

where $p_i = \dfrac{1}{1+\exp\{-(\beta_0 + \beta_1 x_{ji})\}}$

The above equation (5.9) can also be written as:

$$p(\beta_0, \beta_1) \propto H(\beta' x_i)^{-\frac{1}{2}}(1-H(\beta' x_i))^{-\frac{1}{2}} \quad (5.11)$$

5.3.2.4 Joint Noninformative (Uniform) Prior

We consider the joint noninformative Prior of β_0 and β_1 as Uniform Prior:

$$p(\beta_0, \beta_1) \propto 1 \qquad -\infty < \beta_0 < \infty, -\infty < \beta_1 < \infty \quad (5.12)$$

5.3.3 Joint Posterior Distribution

Now the joint posterior distributions for joint noninformative and informative Priors are given in the following sections:

5.3.3.1 Joint Posterior Distribution Using Normal Prior

Now for the joint posterior distribution of β_0 and β_1, we consider the joint Log Likelihood function (5.6) and the joint Normal prior (5.7), then the joint posterior distribution of β_0 and β_1 is:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n}\{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} \exp\left\{-\frac{1}{2b_0}(\beta_0 - a_0)^2\right\} \exp\left\{-\frac{1}{2b_1}(\beta_1 - a_1)^2\right\}$$

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{p_i}{1-p_i} \right) + \log(1-p_i) \right\} \exp\left\{ -\frac{b_1(\beta_0 - a_0)^2 + b_0(\beta_1 - a_1)^2}{2b_0 b_1} \right\} \tag{5.13}$$

By using logistic regression model with intercept and having only one explanatory variable given in equation (5.1) see section (5.2) we can derive the expression:

$$\log(1-p_i) = -\log\left\{1 + \exp(\beta_0 + \beta_1 x_{fi})\right\} \tag{5.14}$$

Then after replacing equation (5.1) and (5.14) the above posterior distribution (5.13) becomes:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \left[y_i(\beta_0 + \beta_1 x_{fi}) - \log\left\{1 + \exp(\beta_0 + \beta_1 x_{fi})\right\} \right]$$

$$\exp\left\{ -\frac{b_1(\beta_0 - a_0)^2 + b_0(\beta_1 - a_1)^2}{2b_0 b_1} \right\}$$

Let us suppose here that $\theta_i = \beta_0 + \beta_1 x_{fi}$ for further simplification also suppose that:

$\phi_0 = (\beta_0 - a_0)^2, \phi_1 = (\beta_1 - a_1)^2$

Then the joint posterior distribution for β_0 and β_1 will be as:

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[y_i \theta_i - \log\left\{1 + \exp(\theta_i)\right\} \right] \exp\left\{ -\frac{b_1 \phi_0 + b_0 \phi_1}{2b_0 b_1} \right\}$$

Let us again suppose that:

$b^* = b_0 b_1$, $\phi_0' = b_1 \phi_0$, $\phi_1' = b_0 \phi_1$

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left(-\frac{1}{2b^*} (\phi_0' + \phi_1') \right) \left\{ y_i \theta_i - \log\left\{1 + \exp(\theta_i)\right\} \right\} \right]$$

Let $\phi^* = \phi_0' + \phi_1'$

$$p(\beta_0,\beta_1 \mid data)=\frac{1}{k}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\right] \tag{5.15}$$

$$-\infty<\beta_0<\infty,\ -\infty<\beta_1<\infty$$

This is the joint posterior distribution of β_0 and β_1, where k is the normalizing constant. Here our main objective is to estimate the unknown parameters. Then for this purpose if we partially differentiate the above posterior equation (5.15) with respect to β_0 and β_1 then equating to zero. So this numerical solution will provide us the posterior estimates (modes), now for this we precede as follows:

Differentiate (5.15) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial\beta_0}=\frac{\partial}{\partial\beta_0}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\right]$$

$$=\sum_{i=1}^{n}\left\{\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{y_i-\frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}-\right.$$

$$\left.\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\exp\left(-\frac{\phi^*}{2b^*}\right)\left(\frac{1}{b_0}(\beta_0-a_0)\right)\right\}$$

Let $\lambda=\frac{1}{b_0}(\beta_0-a_0)$

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial\beta_0}=\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{\left(y_i-\frac{1}{1+\exp(-\theta_i)}\right)-\lambda\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\right\}\right]$$

Now put for maximizing $\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial\beta_0}=0$

$$\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{\left(y_i-\frac{1}{1+\exp(-\theta_i)}\right)-\lambda\left\{y_i\theta_i-\log\left\{1+\exp(\theta_i)\right\}\right\}\right\}\right]=0 \qquad (5.16)$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

Now again differentiate (5.15) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_1}=\frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{y_i\theta_i-\log\left\{1+\exp(\theta_i)\right\}\right\}\right]$$

$$=\sum_{i=1}^{n}\left\{\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{y_i x_{fi}-\frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\dfrac{1}{1+\exp(-\theta_i)}\right)}\right\}-\right.$$

$$\left.\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\exp\left(-\frac{\phi^*}{2b^*}\right)\left(\frac{1}{b_1}(\beta_1-a_1)\right)\right\}$$

Let us suppose that $\lambda^* = \frac{1}{b_1}(\beta_1 - a_1)$

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_0}=\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{\left(y_i-\frac{1}{1+\exp(-\theta_i)}\right)-\lambda^*\left\{y_i\theta_i-\log\left\{1+\exp(\theta_i)\right\}\right\}\right\}\right]$$

Now put for maximizing $\dfrac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_1}=0$

$$\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left\{\left(y_i-\frac{1}{1+\exp(-\theta_i)}\right)-\lambda^*\left\{y_i\theta_i-\log\left\{1+\exp(\theta_i)\right\}\right\}\right\}\right]=0 \qquad (5.17)$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

Posterior modes can be obtained by solving above equations (5.16) and (5.17) numerically.

5.3.3.2 Joint Posterior Distribution Using Haldane Prior

Now for the joint posterior distribution of β_0 and β_1 we consider the joint Log Likelihood function (5.6).and the joint Haldane prior (5.8), then the joint posterior distribution of β_0 and β_1 is found to be:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} p_i^{-1}(1-p_i)^{-1}$$

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i) \right\} p_i^{-1}(1-p_i)^{-1} \qquad (5.18)$$

Replace equation (5.1) see section (5.2) in above expression (5.18) we get:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{fi}) + \log(1-p_i)\} p_i^{-1}(1-p_i)^{-1}$$

Since we know from equation (5.14) $\log(1-p_i) = -\log\{1+\exp(\beta_0 + \beta_1 x_{fi})\}$ that is derived from equation (5.1), now we can derive the expression form equation (5.1) as:

$$p_i^{-1} = 1 + \exp\{-(\beta_0 + \beta x_{fi})\} \;\; \& \;\; (1-p_i)^{-1} = 1 + \exp(\beta_0 + \beta_1 x_{fi}) \qquad (5.19)$$

Now replace (5.14) and (5.19) in (5.18) the joint posterior distribution becomes:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \Big[\{1 + \exp\{-(\beta_0 + \beta x_{fi})\}\} \{1 + \exp(\beta_0 + \beta x_{fi})\}$$

$$y_i(\beta_0 + \beta_1 x_{fi}) - \log\{1 + \exp(\beta_0 + \beta_1 x_{fi})\} \Big]$$

Let us suppose here that:

$$\theta_i = \beta_0 + \beta_1 x_{fi}$$

Then the joint posterior distribution for β_0 and β_1 will be as:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \left[\{2 + \exp(-\theta_i) + \exp(\theta_i)\} \, y_i \theta_i - \log\{1 + \exp(\theta_i)\} \right]$$

$$p(\beta_0,\beta_1 \mid data)=\frac{1}{k}\sum_{i=1}^{n}\left[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\, y_i\theta_i-\log\{1+\exp(\theta_i)\}\right] \tag{5.20}$$

$$-\infty<\beta_0<\infty,\ -\infty<\beta_1<\infty$$

This is the joint posterior distribution of β_0 and β_1, where k is the normalizing constant. Now to obtain the posterior estimates (modes) of the parameters β_0 and β_1 we proceed with partially differentiate the above equation (5.20) with respect to β_0 and β_1 simultaneously and then equating to zero. The numerical solution will provide us the Posterior estimates (modes), for this we precede as follows:

Differentiate (5.20) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial\beta_0}=\frac{\partial}{\partial\beta_0}\sum_{i=1}^{n}\left[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\, y_i\theta_i-\log\{1+\exp(\theta_i)\}\right]$$

$$=\sum_{i=1}^{n}\left\{\{2+\exp(-\theta_i)+\exp(\theta_i)\}\left\{y_i-\frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\exp(\theta_i)\{1+\exp(-\theta_i)\}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}-$$

$$\left.\{1+\exp(\theta_i)\}\exp(-\theta_i)\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{\{2+\exp(-\theta_i)+\exp(\theta_i)\}\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\exp(\theta_i)\{1+\exp(-\theta_i)\}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}-$$

$$\left.\exp(-\theta_i)\{1+\exp(\theta_i)\}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\right\}$$

$$= \sum_{i=1}^{n} \left\{ \{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\{\exp(\theta_i)(1+\exp(-\theta_i)) - \exp(-\theta_i)(1+\exp(\theta_i))\}$$

$$\left. \{y_i\theta_i - \log(1+\exp(\theta_i))\} \right\}$$

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_0} = \sum_{i=1}^{n} \left[\{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \{\exp(\theta_i) - \exp(-\theta_i)\} \right]$$

Now for maximizing put $\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n} \left[\{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \{\exp(\theta_i) - \exp(-\theta_i)\} \right] = 0 \qquad (5.21)$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

Again differentiate (5.20) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} \left[\{2+\exp(-\theta_i)+\exp(\theta_i)\} \, y_i\theta_i - \log\{1+\exp(\theta_i)\} \right]$$

$$= \sum_{i=1}^{n} \left\{ \{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i x_{fi} - \frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2 \left(1 - \frac{1}{1+\exp(-\theta_i)}\right)} \right\} + \right.$$

$$\exp(\theta_i)\{1+\exp(-\theta_i)\} \, x_{fi} \left\{ y_i\theta_i + \log\left(1 - \frac{1}{1+\exp(-\theta_i)}\right) \right\} -$$

$$\left. \{1+\exp(\theta_i)\}\exp(-\theta_i) x_{fi} \left\{ y_i\theta_i + \log\left(1 - \frac{1}{1+\exp(-\theta_i)}\right) \right\} \right\}$$

$$= \sum_{i=1}^{n} \left\{ \{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\exp(\theta_i)\{1+\exp(-\theta_i)\} x_{fi} \left\{ y_i\theta_i + \log\left(\frac{\exp(-\theta_i)}{1+\exp(-\theta_i)} \right) \right\} -$$

$$\left. \{1+\exp(\theta_i)\} \exp(-\theta_i) x_{fi} \left\{ y_i\theta_i + \log\left(\frac{\exp(-\theta_i)}{1+\exp(-\theta_i)} \right) \right\} \right\}$$

$$= \sum_{i=1}^{n} \left\{ \{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\exp(\theta_i)\{1+\exp(-\theta_i)\} x_{fi} \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} -$$

$$\left. \{1+\exp(\theta_i)\} \exp(-\theta_i) x_{fi} \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \right\}$$

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_1} = \sum_{i=1}^{n} \left[\{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. x_{fi} \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \{\exp(\theta_i) - \exp(-\theta_i)\} \right]$$

Now for maximizing put $\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_1} = 0$

$$\sum_{i=1}^{n} \left[\{2+\exp(-\theta_i)+\exp(\theta_i)\} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. x_{fi} \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \{\exp(\theta_i) - \exp(-\theta_i)\} \right] = 0 \qquad (5.22)$$

while $\theta_i = \beta_0 + \beta_1 x_{fi}$

To obtain the Posterior modes solve the equation (5.21) and (5.22) numerically.

5.3.3.3 Joint Posterior Distribution Using Jeffreys Prior

Now for the joint posterior distribution of β_0 and β_1 we consider the joint Log Likelihood function (5.6).and the joint Jeffreys prior (5.10), then the joint posterior distribution of β_0 and β_1 is found to be:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}}$$

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i) \right\} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}} \tag{5.23}$$

Replace equation (5.1) in above expression (5.23) we get:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{fi}) + \log(1-p_i)\} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}} \tag{5.24}$$

Since we know from equation (5.14) $\log(1-p_i) = -\log\{1+\exp(\beta_0 + \beta_1 x_{fi})\}$ that is derived from equation (5.1), now we can derive the expression form equation (5.1) as:

$$p_i^{-\frac{1}{2}} = \sqrt{1+\exp(-\beta_0 - \beta x_{fi})} \;\; \& \;\; (1-p_i)^{-\frac{1}{2}} = \sqrt{1+\exp(\beta_0 + \beta_1 x_{fi})} \tag{5.25}$$

Now replace (5.14) and (5.25) in (5.24) the joint posterior distribution becomes:

$$p(\beta_0, \beta_1 \mid data) \propto \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{fi}) - \log(1+\exp(\beta_0 + \beta_1 x_{fi}))\}$$

$$\sqrt{1+\exp(-\beta_0 - \beta x_{fi})}\sqrt{1+\exp(\beta_0 + \beta_1 x_{fi})}$$

Let us suppose here that, $\theta_i = \beta_0 + \beta_1 x_{fi}$

Then the joint posterior distribution for β_0 and β_1 will be as:

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \{y_i\theta_i - \log(1+\exp(\theta_i))\} \sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)} \tag{5.26}$$

$$-\infty < \beta_0 < \infty, \; -\infty < \beta_1 < \infty$$

This is the joint posterior distribution of β_0 and β_1, where k is the normalizing constant. Here our main objective is to estimate the unknown parameters. Then for this purpose if we partially differentiate the above equation (5.26) with respect to β_0 and β_1 then equating to zero. So this numerical solution will provide us the Posterior estimates (modes), for this we precede as follows.

Differentiate (5.26) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}$$

$$= \sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i - \frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(\theta_i)}} -$$

$$\left.\frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$= \sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i - \frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\{y_i\theta_i - \log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(\theta_i)}} -$$

$$\left.\frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\{y_i\theta_i - \log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$= \sum_{i=1}^{n} \left\{ \sqrt{2+\exp(-\theta_i)+\exp(\theta_i)} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left\{ \frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}}{2\sqrt{1+\exp(\theta_i)}} - \frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}}{2\sqrt{1+\exp(-\theta_i)}} \right\}$$

$$\left. \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \right\}$$

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_0} = \sum_{i=1}^{n} \left\{ \sqrt{2+\exp(-\theta_i)+\exp(\theta_i)} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. \{ y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \left\{ \frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}} \right\} \right\}$$

Now for maximizing put $\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n} \left\{ \sqrt{2+\exp(-\theta_i)+\exp(\theta_i)} \left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} + \right.$$

$$\left. \{ y_i\theta_i - \log(1+\exp(\theta_i))\} \left\{ \frac{\exp(-\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}} \right\} \right\} = 0 \qquad (5.27)$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

Again differentiate (5.26) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} \{ y_i\theta_i - \log(1+\exp(\theta_i))\} \sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}$$

$$= \sum_{i=1}^{n} \left\{ \sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)} \left\{ y_i x_{fi} - \frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2 \left(1 - \frac{1}{1+\exp(-\theta_i)}\right)} \right\} + \right.$$

$$\frac{x_{fi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{x_{fi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\frac{x_{fi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{x_{fi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left\{\frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}}{2\sqrt{1+\exp(\theta_i)}}-\frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$\left.x_{fi}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1\,|\,data)}{\partial\beta_1}=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{fi}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}$$

Now put for maximizing $\frac{\partial p(\beta_0,\beta_1\,|\,data)}{\partial\beta_1}=0$

$$\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left. x_{fi}\left\{y_i\theta_i-\log(1+\exp(\theta_i))\right\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}=0 \qquad (5.28)$$

while $\theta_i = \beta_0 + \beta_1 x_{fi}$

The numerical solution of above equations (5.27) & (5.28) will provide the posterior modes.

5.3.3.4 Joint Posterior distribution Using Uniform Prior

Now using the joint Log likelihood function (5.6).and the joint uniform prior distribution (5.12), then the joint posterior distribution of β_0 and β_1 is found to be:

$$p(\beta_0,\beta_1 \mid data) \propto \sum_{i=1}^{n}\{y_i \log(p_i)+(1-y_i)\log(1-p_i)\}1$$

$$p(\beta_0,\beta_1 \mid data) \propto \sum_{i=1}^{n}\left\{y_i \log\left(\frac{p_i}{1-p_i}\right)+\log(1-p_i)\right\} \qquad (5.29)$$

Now replace equation (5.1) in above expression (5.29) we have:

$$p(\beta_0,\beta_1 \mid data) \propto \sum_{i=1}^{n}\{y_i(\beta_0+\beta_1 x_{fi})+\log(1-p_i)\} \qquad (5.30)$$

Now put equation (5.14) in above equation (5.30) we get:

$$p(\beta_0,\beta_1 \mid data) \propto \sum_{i=1}^{n}\left\{y_i(\beta_0+\beta_1 x_{fi})-\log\left\{1+\exp(\beta_0+\beta_1 x_{fi})\right\}\right\}$$

$$p(\beta_0,\beta_1 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left\{y_i(\beta_0+\beta_1 x_{fi})-\log\left\{1+\exp(\beta_0+\beta_1 x_{fi})\right\}\right\} \qquad (5.31)$$

$$-\infty<\beta_0<\infty,\ -\infty<\beta_1<\infty$$

This is the joint posterior distribution of β_0 and β_1 , where k is the normalizing constant. Here our main objective is to estimate the parameters. Then for this purpose if we partially differentiate the above equation (5.31) with respect to β_0 and β_1 simultaneously and equate it to zero. So this numerical solution will provide us the Posterior estimates (modes), so for this we proceed as follows.

Differentiate (5.31) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_0}=\frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}\left\{y_i(\beta_0+\beta_1 x_{fi})-\log\left\{1+\exp(\beta_0+\beta_1 x_{fi})\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{y_i-\frac{\exp(-(\beta_0+\beta_1 x_{fi}))}{(1+\exp(-\beta_0-\beta_1 x_{fi}))^2\left(1-\dfrac{1}{1+\exp(-\beta_0-\beta_1 x_{fi})}\right)}\right\}$$

$$=\sum_{i=1}^{n}\left\{y_i-\frac{1}{1+\exp(-\beta_0-\beta_1 x_{fi})}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_0}=\sum_{i=1}^{n}\left\{y_i-\frac{1}{1+\exp(-\beta_0-\beta_1 x_{fi})}\right\}$$

Now as we know that $\theta_i=\beta_0+\beta_1 x_{fi}$

Now put for maximizing $\dfrac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_0}=0$

$$\sum_{i=1}^{n}\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}=0 \qquad (5.32)$$

Now differentiate (5.31) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1 \mid data)}{\partial \beta_1}=\frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\left\{y_i(\beta_0+\beta_1 x_{fi})-\log\left\{1+\exp(\beta_0+\beta_1 x_{fi})\right\}\right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi} \exp(-(\beta_0 + \beta_1 x_{fi}))}{(1+\exp(-\beta_0 - \beta_1 x_{fi}))^2 \left(1 - \frac{1}{1+\exp(-(\beta_0 + \beta_1 x_{fi}))}\right)} \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta_0 - \beta_1 x_{fi})} \right\}$$

$$\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_1} = \sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\beta_0 - \beta_1 x_{fi})} \right\}$$

Now put for maximizing $\frac{\partial p(\beta_0, \beta_1 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n} \left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} = 0 \qquad (5.33)$$

while $\theta_i = \beta_0 + \beta_1 x_{fi}$

To obtain the posterior modes, solve the above equations (5.32) and (5.33) numerically.

5.4 Selection of Hyperparameters

Since we know that the prior distribution of parameters β_0 & β_1 are as follows: $\beta_0 \sim N(a_0, b_0)$ and $\beta_1 \sim N(a_1, b_1)$, our main objective here is to find the values of these hyperparameters while a_0 & a_1 are the means of prior distributions and b_0 & b_1 are the variances of the prior distributions. We have suggested a range of values of hyperparameters by observing the variation in regression coefficients and also the variable of interest and select the values with minimum standard error, the values are given as follows:

Table 5.1

Posterior Estimates Using Different Values of Hyperparameters

Hyperparameters				Posterior Mode		Posterior Mean		Standard Error	
Mean a_0	*Variance* b_0	*Mean* a_1	*Variance* b_1	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
0	1	0	1	−7.0703	2.0535	−8.0213	2.3391	2.7645	0.9224
15	5.5	10.5	2.5	−13.9789	5.8436	−50.8567	14.7011	17.9522	4.8567
14	5	9.5	2	−12.7249	4.3121	−47.0578	13.6067	15.6859	4.5366
13	4.5	8.5	1.5	−11.1784	3.6599	−43.3829	12.5478	14.4010	4.1859
12	4	8	1	−10.1870	2.6710	−39.7078	11.4888	13.9660	3.8295
11	3.5	7.5	0.90	−9.2572	2.1710	−36.1941	10.4761	12.8647	3.4923
10	3	6.5	0.80	−7.9225	2.0893	−32.9411	9.5381	10.9812	3.1594
9	2.90	6.25	0.70	−7.5597	2.0194	−29.6960	8.6023	9.8987	2.8474
8.5	2.80	6.10	0.65	−7.4635	1.9581	−26.1416	8.1540	8.7114	2.7189
8.25	2.70	5.95	0.60	−7.3427	1.8923	−23.3924	7.7993	7.7975	2.6077
8.10	2.60	5.75	0.55	−7.2664	1.8583	−21.9642	7.1843	7.3214	2.3948
7.95	2.50	5.50	0.50	−7.0985	1.8295	−18.5474	6.6932	6.1825	2.2311
7.75	2.25	5.40	0.45	−6.9371	1.8023	−15.9575	5.5238	5.3090	1.8413
7.60	2.10	5.30	0.40	−6.8525	1.7898	−13.2650	4.3238	4.4217	1.4413
7.50	2	5	0.38	−6.7569	1.7724	−11.5244	3.9387	3.8416	1.3529
7.40	1.95	4.90	0.36	−6.6391	1.7699	−9.9825	3.2423	3.2395	1.0509
7.30	1.90	4.80	0.35	−6.5237	1.7618	−8.7012	2.6711	2.9506	0.9305
7.20	1.85	4.75	0.34	−6.4522	1.7598	−7.9436	2.3788	2.6579	0.8669
7.10	1.80	4.70	0.33	−6.4286	1.7513	−7.5020	2.0996	2.5115	0.8293

These are the values for hyperparameters for informative priors which are Normal priors for each parameter that is considered independent. Where *mean* $= a_0$ and *variance* $= b_0$ for the prior distribution of β_0 while *mean* $= a_1$ and *variance* $= b_1$ for the prior distribution of β_1.

We suggest different values for the hyperparameters and find the values of parameters. So finally we decided to select the values of hyperparameters as *mean* = *7.10* and *variance* = *1.80* for the prior distribution of β_0 and *mean* = *4.70* and *variance* = *0.33* they have the minimum standard error.

5.5 Bayesian Analysis with Informative and Noninformative Priors

In this section we will carry out the Bayesian analysis with informative priors and noninformative priors. The analysis is based on the posterior distributions that are derived in previous sections:

5.5.1 Bayesian Analysis Using Joint Normal Prior

In this section we deal with the Bayesian analysis of logistic regression model with intercept and one explanatory variable by using informative (Joint Normal) prior. We use the following joint posterior distribution for the parameters β_0 & β_1 derived in section (5.3.3.1), see equation (5.15):

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right]$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty$$

where k is the normalizing constant:

Programs in SAS package have been designed to show the graph of marginal densities of the parameters β_0 & β_1 by using the data set given in Table 4.1. Similar program is given in Appendix IV.

Graph of Posterior Marginal Densities using Normal Prior

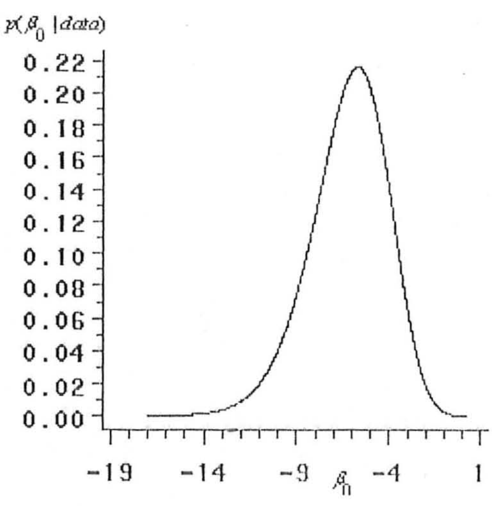

Figure 1(a)

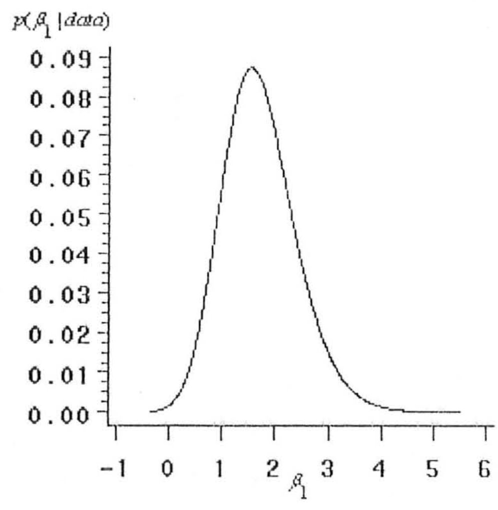

Figure 1(b)

5.5.1.1 Posterior Estimates

For further analysis we have designed programs in SAS package, similar program is given in Appendix II and also a similar program for is given in appendix IV to obtain the value of standard error while using the data set given in Table 4.1 and hyperparameters that are obtain in section (5.4.2). We have used Marquart method to obtain the posterior modes while Quadrature method is used to obtain posterior mean and standard error.

Table 5.2

Posterior Estimates Using Joint Normal Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	−7.5010	−6.5485	2.5115		−0.3797
$\hat{\beta}_1$	2.0996	1.8093	0.8293	6.1062	0.3500

Here we have observed that the posterior mode for β_0 is greater then the posterior mean of β_0 which indicates that the distribution of this parameter is negatively skewed. This is also indicated in graph see figure 1(a). How much it is skewed we have calculated the coefficient of skewness. We have also observed that the posterior mean of β_1 is greater then the posterior mode of β_1 which shows that the distribution of parameter is positively skewed, that is also indicated in figure 1(b), which can also be seen by the coefficient of skewness given in Table 5.2. It is observed that the odds ratio is greater then 1 which indicates that the variable fibrinogen is more likely to occur, so the odds ratio is high for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h. So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately 6.1062 unit increases in the level of ESR. That is too high as compared to the results given in previous chapter. So it can be concluded that the strength of relationship between the probability of an ESR reading greater than 20 mm/h and the level of protein plasma (fibrinogen) is high.

5.5.1.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the joint posterior distribution upon the parameter i.e.

We test the following hypotheses:

$H_0 : \beta_1 \le 0$ Versus $H_1 : \beta_1 > 0$

The posterior probability for H_0 is:

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

Now the posterior probability using informative prior is:

$$p_0 = \int_{-5}^{0} \int_{-15}^{15} \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left(-\frac{\phi^*}{2b^*} \right) \{ y_i \theta_i - \log\{1 + \exp(\theta_i)\}\} \right] d\beta_0 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_0 = 0.001873$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.19% chance to accept H_0 so accept H_1 with high probability and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein rises in the blood plasma. That is also indicated in the result of odds ratio.

5.5.2 Bayesian Analysis Using Joint Haldane Prior

In this section we will present the Bayesian analysis of binary logistic regression model with intercept and one explanatory variable by using noninformative (Joint Haldane) prior. We use the following joint posterior distribution for the parameters β_0 & β_1 derived in section (5.3.3.2) see equation (5.20):

$$p(\beta_0,\beta_1 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\, y_i\theta_i - \log\{1+\exp(\theta_i)\}\right]$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty$$

where k is the normalizing constant:

Programs in SAS package have been designed; similar program is given in appendix IV to show the graphs of marginal densities of the parameters β_0 & β_1 by using the data set given in Table 4.1.

Graph Posterior Marginal Densities using Haldane Prior

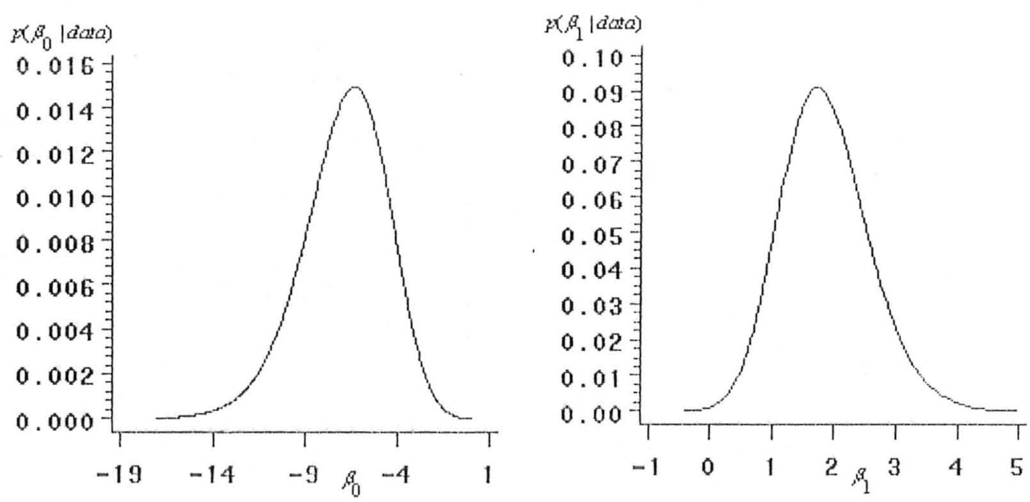

Figure 2(a) Figure 2(b)

5.5.2.1 Posterior Estimates

For further analysis we have designed program in SAS package, similar program is given in appendix II and use the data set given in Table 4.1. We have used Marquart method to obtain the posterior modes while Quadrature method is used to obtain posterior mean and standard error.

Table 5.3

Posterior Estimates Using Joint Haldane Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	−7.6910	−6.6925	2.6015		−0.3838
$\hat{\beta}_1$	2.1872	1.8885	0.8515	6.6094	0.3508

Here we have observed that the posterior mode for β_0 is greater then the posterior mean of β_0 which indicates that the distribution of parameter is negatively skewed, as indicated in graph figure 2(a), which can also be checked by coefficient of skewness given in Table 5.3. We have also observed that the posterior mean of β_1 is greater then the posterior mode of β_1 which shows that the distribution of this parameter is positively skewed, this is also shown in graph figure 2(b), how much it is skewed can be seen by the coefficient of skewness. It is also observed that the odds ratio is greater then 1 which indicates that the variable fibrinogen is more likely to occur, So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately 6.6094 unit increases in the level of ESR. So there is no difference in the value of odds ratio as given in the previous model. So it can be concluded that the strength of relationship between the probability of an ESR reading greater than 20 mm/h and the level of protein plasma (fibrinogen) is also high for this model.

5.5.2.2 Bayesian Hypothesis Testing

The posterior probability of hypothesis given in section (5.5.1.2) using Haldane prior is:

$$p_0 = \int_{-5}^{0}\int_{-16}^{16}\sum_{i=1}^{n}\left[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\, y_i\theta_i - \log\{1+\exp(\theta_i)\}\right]d\beta_0 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.002014$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.20% chance to accept H_0 so we accept H_1 with high probability and conclude that β_1 is positive and playing a significant role to effect the ESR if this protein rises in the blood plasma. This result also gives the evidence in favor of odds ratio that also indicate the same results.

5.5.3 Bayesian Analysis Using Joint Jeffreys Prior

In this section we will present the Bayesian analysis of binary logistic regression model with intercept and one explanatory variable by using noninformative (Joint Jeffreys) prior. We use the following joint posterior distribution for the parameters β_0 & β_1 derived in section (5.3.3.3) see equation (5.26):

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty$$

Where k is the normalizing constant:

Programs in SAS package have been designed; similar program is given in appendix IV to show the graph of marginal densities of the parameters β_0 & β_1 by using the data set in Table 4.1.

Graph of Posterior Marginal Densities using Jeffreys Prior

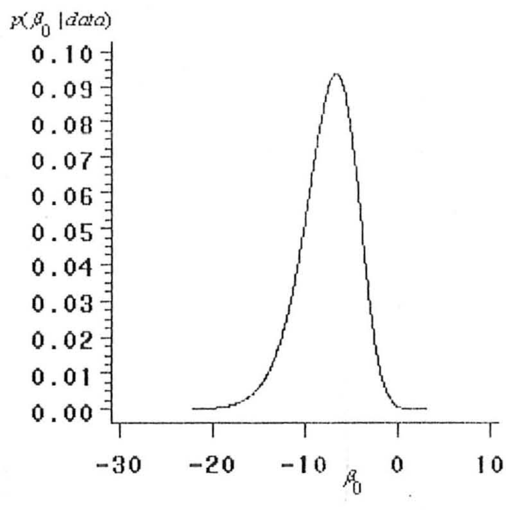

Figure 3(a)

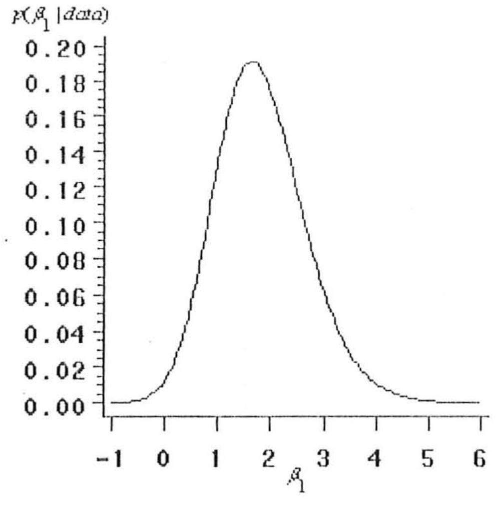

Figure 3(b)

5.5.3.1 Posterior Estimates

For further analysis we have designed programs in SAS package, given in appendix II and use the data set given in Table 4.1. We have used Marquart method to obtain the posterior modes, while Quadrature method is used to obtain posterior mean and standard error.

Table 5.4

Posterior Estimates Using Joint Jeffreys Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	−7.9124	−6.8575	2.7284		−0.3866
$\hat{\beta}_1$	2.2519	1.9236	0.9115	6.8456	0.3602

We can observed that posterior mode for β_0 is greater then the posterior mean of β_0 which indicates that the distribution of this parameter is negatively skewed which is also observed through the presentation of graph in figure 3(a) and also the coefficient of skewness given in Table 5.3. We have also observed that the posterior mean of β_1 is greater then the posterior mode of β_1 which shows that the distribution of this parameter is positively skewed how much it is skewed can be seen by the coefficient of skewness given in Table 5.3 and the graph in figure 3(b) also indicates the same skewness. It is also observed that the odds ratio is greater then 1 which indicates that the variable fibrinogen is more likely to occur, So the strength of relationship between the probability of an ESR reading greater than 20 mm/h and the level of protein plasma (fibrinogen) is also high for this model, as one unit increase in the level of protein plasma (fibrinogen) increases the ESR by 6.8456 units. That is also same as for the other models but with slight difference.

5.5.3.2 Bayesian Hypothesis Testing

The posterior probability of hypotheses given in section (5.5.1.2) using Jeffreys prior is:

$$p_0 = \int_{-5}^{0}\int_{-20}^{20} \frac{1}{k} \sum_{i=1}^{n} \{y_i\theta_i - \log(1+\exp(\theta_i))\} \sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\, d\beta_0 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_0 = 0.002157$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.22% chance to accept H_0 so we accept H_1 with high probability and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma. This evidence also favors the results given in Table 5.4.

5.5.4 Bayesian Analysis Using Joint Uniform Prior

In this section we will present the Bayesian analysis of binary logistic regression model with intercept and one explanatory variable by using noninformative (Joint Uniform) prior. We use the following joint posterior distribution for the parameters β_0 & β_1 derived in section (5.3.3.4) see equation (5.31).

$$p(\beta_0, \beta_1 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left\{y_i(\beta_0 + \beta_1 x_{fi}) - \log\left\{1 + \exp(\beta_0 + \beta_1 x_{fi})\right\}\right\}$$

$$-\infty < \beta_0 < \infty, \quad -\infty < \beta_1 < \infty$$

Where k is the normalizing constant:

Programs in SAS package has been designed, similar program is given in appendix IV to show the graph of marginal densities of the parameters β_0 & β_1 by using the data set in Table 4.1.

Graph of Posterior Marginal Densities using Uniform Prior

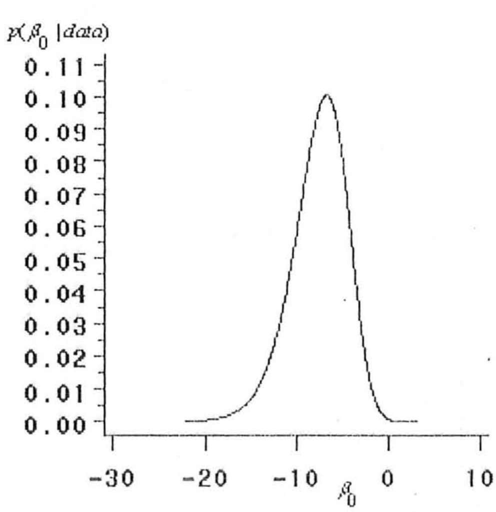

Figure 4(a)

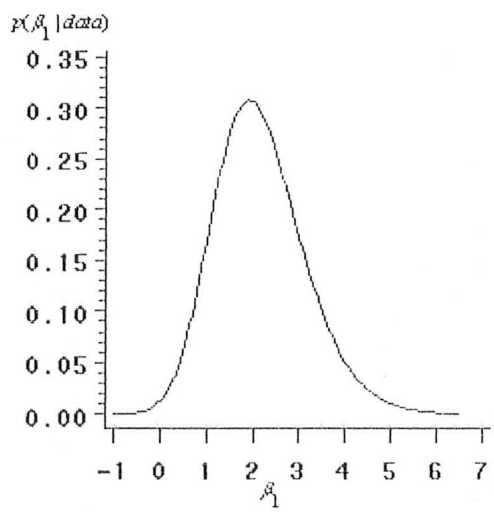

Figure 4(b)

5.5.4.1 Posterior Estimates

For further analysis we have designed program in SAS package, similar program is given in appendix II and use the data set given in Table 4.1. We have used Marquart method to obtain the posterior mode while Quadrature method is used to obtain posterior mean and standard error by using the Posterior distribution:

Table 5.5

Posterior Estimates Using Joint Uniform Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	-8.0574	-6.9638	2.7770		-0.3938
$\hat{\beta}_1$	2.2882	1.9499	0.9169	7.0280	0.3690

Here we have seen that the posterior mode for β_0 is greater then the posterior mean of β_0 which indicates that the distribution of this parameter is negatively skewed how much it is skewed we have calculated the coefficient of skewness. The graph of this parameter given in figure 4(a) also indicates the same pattern of skewness. We have also observed that the posterior mean of β_1 is greater then the posterior mode of β_1 which shows that the distribution of parameter is positively skewed how much it is skewed we have calculated the coefficient of skewness given in Table 5.5 and the graph of this parameter given in figure 4(b) shows a positively skewed shape. Here the odds ratio is also greater then 1 which indicates that the variable fibrinogen is more likely to occur, So the strength of relationship between the probability of an ESR reading greater than 20 mm/h and the level of protein plasma (fibrinogen) is also high for this model, as one unit increase in the level of protein plasma (fibrinogen) increases the ESR by 7.0280 units. The results are different but with slight change, that does not effect the interpretation of odds ratio that we did for previous models.

5.5.4.2 Bayesian Hypothesis Testing

Now to test the hypothesis that whether fibrinogen level in the blood play any significance role to increase ESR as if the level of certain proteins in the blood plasma rise. The posterior probability for H_0 using Joint Uniform prior is calculated as:

The posterior probability of hypotheses given in section (5.5.1.2) using (Joint Uniform) prior is:

$$p_0 = \int_{-6}^{0}\int_{-20}^{20} \frac{1}{k} \sum_{i=1}^{n} \left\{ y_i(\beta_0 + \beta_1 x_{fi}) - \log\left\{1 + \exp(\beta_0 + \beta_1 x_{fi})\right\}\right\} d\beta_0 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi}$

A program has been designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.002417$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.24% chance to accept H_0 so we accept H_1 with high probability but little low as for the previous models and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma. That is also the case in the results given in Table 5.5

5.6 Classical Regression Analysis

For the comparison purpose now we take the classical estimates and test the hypotheses. For this we have simply run the logistic regression without intercept model. Now the classical estimate and hypothesis testing is given in following section:

5.6.1 Classical Estimates

Using the data given in Table 4.1 and having run the logistic regression we obtain:

Table 5.6

Output of Logistic Regression Using Classical Approach

Coefficient	Classical Estimate	Standard Error	Z-Statistic	P-Value	Odds Ratio
$\hat{\beta}_0$	−6.9640	2.7770	−2.5100	0.0120	
$\hat{\beta}_1$	1.9499	0.9169	2.1300	0.0330	7.0300

5.6.2 Classical way of Hypothesis Testing

We have the logistic regression model as:

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

Hypothesis

$H_0 : \beta_1 \geq 0$ Versus $H_1 : \beta_1 < 0$

Since $\beta_1 = 1.9499$ and standard error of β is 0.9199 then the value of Wald t-statistic is:

$t = 2.1300$

Since the p-value from this regression is 0.0300, it indicates that we accept H_0 up to 3% level of significance and do not accept H_0 at any other level of significance. So it can be concluded that fibrinogen is playing significant role at 5% level of significance and it effect the ESR if this protein rises in the blood plasma.

5.7 Comparison of Bayesian and Classical Logistic Regression Analysis

Now as a summery we present the results of Logistic regression model with intercept having one explanatory variable, we have obtained by using Bayesian and Classical techniques and make comparison between these two, the results are presented in table (5.7) are the Bayesian results obtained by using different priors these results can be compared with the results given in table (5.6):

Table 5.7

Posterior Estimates for With Intercept Logistic Regression Model

Coefficient	Posterior Estimates	Noninformative Prior			Informative prior
		Uniform Prior	Jeffreys prior	Haldane prior	
$\hat{\beta}_0$	Posterior Mode	-6.9638	-6.8575	-6.6925	-6.5485
	Posterior Mean	-8.0574	-7.9124	-7.6910	-7.5010
	Standard Error	2.7770	2.7284	2.6015	2.5115
	SK_p	-0.3938	-0.3866	-0.3838	-0.3797
$\hat{\beta}_1$	Posterior Mode	1.9499	1.9236	1.8885	1.8093
	Posterior Mean	2.2882	2.2519	2.1872	2.0996
	Odds Ratio	7.0280	6.8456	6.6094	6.1062
	Standard Error	0.9169	0.9115	0.8515	0.8293
	SK_p	0.3690	0.3602	0.3508	0.3500

The results found by using Classical logistic regression and in Bayesian logistic regression with Uniform prior are approximately same in all respects i.e. the coefficients, p-value and odds ratio. Here odds ratio interpreted as the approximated change in the risk of disease for every one unit increase in the amount of fibrinogen. So the results are much improved with Haldane and informative prior as compared to uniform and Jeffreys.

Chapter 6

Bayesian Inference of Logistic Regression Model with Two Explanatory Variables

6.1 Introduction

In this chapter, we present the Bayesian analysis of logistic regression model with two explanatory variables for response binary variable, under informative and noninformative priors. Section 2 consist the data set used for the analysis of binary logistic regression model with two explanatory variables. The derivation of posterior distribution using informative and noninformative prior is given in section 3.In section 4 for informative prior, we set a range of hyperparameters and select the hyperparameters with minimum standard error. The idea of selecting hyperparameters is taken from Bian (1997); they assume Normal & Student-t priors for regression coefficients with mean zero and decide about the posterior distribution at different values of variances. But we have suggested a range of values for all hyperparameters. Section 5 provides the Bayesian analysis with informative and noninformative priors, which include the graphs, posterior estimates (modes & means), standard errors and testing the hypotheses concerning parameters. Section 6 presents the classical analysis of logistic regression model with two explanatory variables and also the hypothesis testing for the significance of regression coefficients. In the last section 7 the comparison of classical and Bayesian results and their interpretation are discussed.

6.2 Data set to be used in Bayesian Logistic Regression Analysis

The data set for Bayesian analysis of Binary Logistic Regression 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. The data set of 32 observations is given 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

Table 6.1: Data

Serial No.	ESR (mm/h)	Fibrinogen (gm/l)	Y-globulin (gm/l)	Serial No.	ESR (mm/h)	Fibrinogen (gm/l)	Y-globulin (gm/l)
	y_i	x_{fi}	x_{gi}		y_i	x_{fi}	x_{gi}
1	0	2.52	38	17	1	3.53	46
2	0	2.56	31	18	0	2.68	34
3	0	2.19	33	19	0	2.60	38
4	0	2.18	31	20	0	2.23	37
5	0	3.41	37	21	0	2.88	30
6	0	2.46	36	22	0	2.65	46
7	0	3.22	38	23	1	2.09	44
8	0	2.21	37	24	0	2.28	36
9	0	3.15	39	25	0	2.67	39
10	0	2.60	41	26	0	2.29	31
11	0	2.29	36	27	0	2.15	31
12	0	2.35	29	28	0	2.54	28
13	1	5.06	37	29	1	3.93	32
14	1	3.34	32	30	0	3.34	30
15	1	2.38	37	31	0	2.99	36
16	1	3.15	36	32	0	3.32	35

6.3 Joint Posterior Distribution for the Parameters of the Logistic Regression Model

Here we will consider the binary logistic regression model with two explanatory variables as:

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi} \quad (6.1)$$

Here β_0 is the intercept while β_1 and β_2 are the slope coefficients for the explanatory variables fibrinogen and Y-globulin respectively. The above logistic regression model can also be represented as:

$$p_i = p_r(y_i = 1) = \frac{1}{1+\exp\{-(\beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi})\}} \quad (6.2)$$

Then the joint Posterior distribution of the parameters β_0, β_1 and β_2 are defined as:

$$p(\beta_0, \beta_1, \beta_2 \mid data) \propto l(\beta_0, \beta_1, \beta_2 \mid data) \times p(\beta_0, \beta_1, \beta_2) \quad (6.3)$$

Here $p(\beta_0, \beta_1, \beta_2 \mid data)$ is the joint posterior distribution while $l(\beta_0, \beta_1, \beta_2 \mid data)$ is the joint log likelihood function and $p(\beta_0, \beta_1, \beta_2)$ is the joint prior distribution for β_0, β_1 and β_2. Here we considered that the explanatory variables are independent of each other. Now we need to decide upon the joint prior distribution and the log likelihood function, we also considered here Log likelihood instead of simple likelihood function just for the ease in calculation.

6.3.1 Joint Likelihood Function

The joint likelihood of the i^{th} observation is its probability density function as a function of the two parameters β_0, β_1 and β_2 where (y_i, x_{fi}, x_{gi}) are fixed at the observed values. The observations are all independent, now for the given case we precede as follows.

Let y_i be the response variable that is binary in nature i.e. it takes only two values 0 and 1 for 'n' observations. Since the analysis of binary response variable in classical approach, the Maximum Likelihood Method (MLE) is used to estimate the unknown parameters of the Binary Logistic Regression (BLR) Model. However the estimates based on the classical approach are not accurate when the sample size is small. In this situation Bayesian approach provides better and accurate results. Then if y_i is the response variable while x_{fi} and x_{gi} are the explanatory variables that can be either qualitative or quantitative in nature while p_i is the probability of success corresponding to the i^th^ observation then the joint likelihood function can be presented as:

$$L(\beta' \mid data) = \prod_{i=1}^{n} \left\{ p_i^{y_i} (1-p_i)^{1-y_i} \right\} \tag{6.4}$$

Now if $p_i = H(\beta' x_i)$ while $\beta' = (\beta_0, \beta_1, \beta_2)$ then the joint likelihood function could be written as:

$$L(\beta' \mid data) = \prod_{i=1}^{n} \left\{ H(\beta' x_i)^{y_i} (1 - H(\beta' x_i))^{1-y_i} \right\} \tag{6.5}$$

As we know that in logistic regression $p_i = H(\beta' X)$ while β' is the vector of regression coefficients and X the set of explanatory variables. While H is the link function that is logistic in our case and we will use this link through out our study to obtain the posterior estimates.

Taking log on both sides of above equation (6.5) we get the Joint Log likelihood function becomes:

$$l(\beta' \mid data) = \sum_{i=1}^{n} \left\{ y_i \log H(\beta' x_i) + (1 - y_i) \log(1 - H(\beta' x_i)) \right\} \tag{6.6}$$

Sine we know that $p_i = H(\beta' x_i)$ then for further consideration we can write the above log likelihood function as:

$$l(\beta_0, \beta_1, \beta_2 \mid data) = \sum_{i=1}^{n} \{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \} \qquad (6.7)$$

where p_i is the probability of success for i[th] observation in data set and be represented in the logistic regression model as:

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

6.3.2 Joint Prior Distribution

We consider the joint noninformative and informative priors of β_0, β_1 and β_2 in the following sections:

6.3.2.1 Joint Informative (Normal) Prior

Here we consider the independent normal priors for each parameter. The joint Informative prior of three parameters is the product of the three individual priors:

$$p(\beta_0, \beta_1, \beta_2) = p(\beta_0) p(\beta_1) p(\beta_2)$$

While $\beta_0 \sim N(a_0, b_0)$, $\beta_1 \sim N(a_1, b_1)$ and $\beta_2 \sim N(a_2, b_2)$ here a_0, a_1 and a_2 are means while b_0, b_1 and b_2 are the variances. Therefore

$$p(\beta_0, \beta_1, \beta_2) \propto \exp\left\{ -\frac{1}{2b_0} (\beta_0 - a_0)^2 \right\} \exp\left\{ -\frac{1}{2b_1} (\beta_1 - a_1)^2 \right\} \exp\left\{ -\frac{1}{2b_2} (\beta_2 - a_2)^2 \right\}$$

$$p(\beta_0, \beta_1, \beta_2) \propto \exp\left\{ -\frac{b_1 b_2 (\beta_0 - a_0)^2 + b_0 b_2 (\beta_1 - a_1)^2 + b_0 b_1 (\beta_2 - a_2)^2}{2 b_0 b_1 b_2} \right\} \qquad (6.8)$$

$$-\infty < \beta_0 < \infty, \quad -\infty < \beta_1 < \infty, \quad -\infty < \beta_2 < \infty$$

6.3.2.2 Joint Noninformative (Haldane) prior

The joint noninformative (Haldane) prior by using the log likelihood function given in (6.7) is derived as:

$$p_H(\beta_0,\beta_1,\beta_2) \propto p_i^{-1}(1-p_i)^{-1} \qquad (6.9)$$

Where $p_i = \dfrac{1}{1+\exp\{-(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})\}}$

The above equation (6.9) can also be written as:

$$p_H(\beta_0,\beta_1,\beta_2) \propto H(\beta' x_i)^{-1}(1-H(\beta' x_i))^{-1} \qquad (6.10)$$

6.3.2.3 Joint Noninformative (Jeffreys) Prior

The joint noninformative (Jeffreys) prior by using the log likelihood function given in (6.7) is derived as:

$$p_J(\beta_0,\beta_1,\beta_2) \propto \sqrt{\det\left|I(\beta_0,\beta_1,\beta_2)\right|}$$

Then $p_J(\beta_0,\beta_1,\beta_2) \propto p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}}$ (6.11)

Where $p_i = \dfrac{1}{1+\exp\{-(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})\}}$

The above equation (6.11) can also be written as:

$$p_J(\beta_0,\beta_1,\beta_2) \propto H(\beta' x_i)^{-\frac{1}{2}}(1-H(\beta' x_i))^{-\frac{1}{2}} \qquad (6.12)$$

6.3.2.4 Joint Noninformative (Uniform) Prior

We consider the joint noninformative Prior of β_0, β_1 and β_2 as Uniform Prior which can be represented as:

$$p(\beta_0,\beta_1,\beta_2) \propto 1 \qquad -\infty<\beta_0<\infty, -\infty<\beta_1<\infty, -\infty<\beta_2<\infty \qquad (6.13)$$

6.3.3 Joint Posterior Distribution

Now the joint posterior distributions for joint informative and noninformative Priors are given in the following sections:

6.3.3.1 Joint Posterior Distribution Using Normal Prior

Now for the joint posterior distribution of β_0, β_1 and β_2 we consider the joint Log Likelihood function given in equation (6.7) and the joint Normal prior given in equation (6.8), then the joint posterior distribution of β_0, β_1 and β_2 is found to be:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} \exp\left\{-\frac{1}{2b_0}(\beta_0 - a_0)^2\right\}$$

$$\exp\left\{-\frac{1}{2b_1}(\beta_1 - a_1)^2\right\} \exp\left\{-\frac{1}{2b_2}(\beta_2 - a_2)^2\right\}$$

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\left\{y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right\}$$

$$\exp\left\{-\frac{b_1b_2(\beta_0 - a_0)^2 + b_0b_2(\beta_1 - a_1)^2 + b_0b_1(\beta_2 - a_2)^2}{2b_0b_1b_2}\right\} \qquad (6.14)$$

by using logistic regression model with two explanatory given in equation (6.1) see section (6.3) we can derive the expression:

$$\log(1-p_i) = -\log\{1 + \exp(\beta_0 + \beta_1 x_{ji} + \beta_2 x_{gi})\} \qquad (6.15)$$

Then after replacing equation (6.1) and (6.15) the above posterior distribution (6.14) becomes:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i(\beta_0 + \beta_1 x_{ji} + \beta_2 x_{gi}) - \log\{1 + \exp(\beta_0 + \beta_1 x_{ji} + \beta_2 x_{gi})\}\}$$

$$\exp\left\{-\frac{b_1b_2(\beta_0 - a_0)^2 + b_0b_2(\beta_1 - a_1)^2 + b_0b_1(\beta_2 - a_2)^2}{2b_0b_1b_2}\right\}$$

For further simplification let us also consider that $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$ and suppose that:

$\phi_0 = (\beta_0 - a_0)^2$, $\phi_1 = (\beta_1 - a_1)^2$ and $\phi_2 = (\beta_2 - a_2)^2$

Now the simplified joint posterior distribution of β_0, β_1 and β_2 can be rewritten as:

$$p(\beta_0, \beta_1, \beta_2 \mid data) \propto \sum_{i=1}^{n} \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \exp\left\{-\frac{b_1 b_2 \phi_0 + b_0 b_2 \phi_1 + b_0 b_1 \phi_2}{2 b_0 b_1 b_2}\right\}$$

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left\{ -\frac{1}{2}\left(\frac{\phi_0}{b_0} + \frac{\phi_1}{b_1} + \frac{\phi_2}{b_2} \right) \right\} \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \right]$$

Let us also suppose that:

$b^* = b_0 b_1 b_2$, $\phi_0' = b_0 b_1 \phi_0$, $\phi_1' = b_0 b_2 \phi_1$ and $\phi_2' = b_0 b_1 \phi_2$

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left\{ -\frac{1}{2b^*}\left(\phi_0' + \phi_1' + \phi_2' \right) \right\} \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \right]$$

Let $\phi^* = \phi_0' + \phi_1' + \phi_2'$

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left(-\frac{\phi^*}{2b^*} \right) \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\} \right] \qquad (6.16)$$

$$-\infty < \beta_0 < \infty, \; -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

This is the joint posterior distribution of β_0, β_1 and β_2, where k is the normalizing constant. Here our main objective is to estimate the unknown parameters. Then for this purpose if we partially differentiate the above posterior equation (6.16) with respect to β_0, β_1 and β_2 then equating to zero. So this numerical solution will provide us the posterior estimates (modes), now for this we precede as follows:

Differentiate (6.16) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right]$$

$$= \sum_{i=1}^{n}\left\{\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left\{y_i - \frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\dfrac{1}{1+\exp(-\theta_i)}\right)}\right\} - \right.$$

$$\left.\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left(\frac{1}{b_0}(\beta_0 - a_0)\right)\right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = \sum_{i=1}^{n}\left\{\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i - \frac{1}{1+\exp(-\theta_i)}\right) - \right.$$

$$\left.\frac{1}{b_0}(\beta_0 - a_0)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i - \frac{1}{1+\exp(-\theta_i)}\right) - \gamma\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right] = 0 \qquad (6.17)$$

While $\gamma = \frac{1}{b_0}(\beta_0 - a_0)$

Again differentiate (6.16) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right]$$

$$= \sum_{i=1}^{n}\left\{\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left\{y_i x_{fi} - \frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\dfrac{1}{1+\exp(-\theta_i)}\right)}\right\} - \right.$$

$$\left\{y_i\theta_i + \log\left(1 - \frac{1}{1+\exp(-\theta_i)}\right)\right\}\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left(\frac{1}{b_1}(\beta_1 - a_1)\right)\Bigg\}$$

$$\frac{\partial p(\beta_0, \beta_1, \beta_2 \mid data)}{\partial \beta_1} = \sum_{i=1}^{n}\left\{\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)}\right) - \right.$$

$$\left. \frac{1}{b_1}(\beta_1 - a_1)\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0, \beta_1, \beta_2 \mid data)}{\partial \beta_1} = 0$

$$\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)}\right) - \gamma'\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right] = 0 \qquad (6.18)$$

where $\gamma' = \frac{1}{b_1}(\beta_1 - a_1)$

Again differentiate (6.16) with respect to β_2 we obtain:

$$\frac{\partial p(\beta_0, \beta_1, \beta_2 \mid data)}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right]$$

$$= \sum_{i=1}^{n}\left\{\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left\{y_i x_{gi} - \frac{x_{gi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1 - \frac{1}{1+\exp(-\theta_i)}\right)}\right\} - \right.$$

$$\left\{y_i\theta_i + \log\left(1 - \frac{1}{1+\exp(-\theta_i)}\right)\right\}\exp\left\{-\frac{1}{2b_0}\phi_0 - \frac{1}{2b_1}\phi_1 - \frac{1}{2b_2}\phi_2\right\}\left(\frac{1}{b_2}(\beta_2 - a_2)\right)\Bigg\}$$

$$\frac{\partial p(\beta_0, \beta_1, \beta_2 \mid data)}{\partial \beta_2} = \sum_{i=1}^{n}\left\{\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i x_{gi} - \frac{x_{gi}}{1+\exp(-\theta_i)}\right) - \right.$$

$$\left. \frac{1}{b_2}(\beta_2 - a_2)\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_2} = 0$

$$\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\left(y_i x_{gi} - \frac{x_{gi}}{1+\exp(-\theta_i)}\right) - \gamma''\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right] = 0 \tag{6.19}$$

while $\gamma'' = \frac{1}{b_2}(\beta_2 - a_2)$

now solving numerically the above equations (6.17), (6.18) and (6.19) the posterior modes of β_0, β_1 and β_2 can be obtain.

6.3.3.2 Joint Posterior Distribution Using Haldane Prior

Now for the joint posterior distribution of β_0, β_1 and β_2 we consider the joint Log Likelihood function (6.7).and the joint Haldane prior (6.9), then the joint posterior distribution of β_0, β_1 and β_2 is found to be:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i \log(p_i) + (1-y_i)\log(1-p_i)\}p_i^{-1}(1-p_i)^{-1}$$

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\left\{y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right\}p_i^{-1}(1-p_i)^{-1} \tag{6.20}$$

By using logistic regression model with two explanatory given in equation (6.1) see section (6.3) and the expression derived from (6.1) that is equation (6.15):

Then after replacing equation (6.1) and (6.15) the above posterior distribution (6.20) becomes:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi}) - \log\{1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})\}\}p_i^{-1}(1-p_i)^{-1}$$

Now we will also use equation (6.1) to derive expressions given as:

$$p_i^{-1} = 1+\exp\{-(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi})\} \ \ \& \ \ (1-p_i)^{-1} = 1+\exp(\beta_0+\beta_1 x_{fi}+\beta_2 x_{gi}) \tag{6.21}$$

As we know we have already suppose that $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$ then the joint Posterior distribution for β_0, β_1 and β_2 will become as:

$$p(\beta_0, \beta_1, \beta_2 \mid data) \propto \sum_{i=1}^{n} \{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\{1+\exp(-\theta_i)\}\{1+\exp(\theta_i)\}$$

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i - \log(1+\exp(\theta_i)\}] \qquad (6.22)$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

The equation (6.22) is the joint posterior distribution of β_0, β_1 and β_2, where k is the normalizing constant. Now to estimate the known parameters β_0, β_1 and β_2, partially differentiate the above equation (6.22) with respect to β_0, β_1 and β_2 simultaneously and then equating to zero. So this numerical solution will provide us the Posterior estimates (modes), for this we precede as follows:

Differentiate (6.22) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0, \beta_1, \beta_2 \mid data)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i - \log(1+\exp(\theta_i)\}]$$

$$= \sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_i - \frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\dfrac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\} -$$

$$\left.(1+\exp(\theta_i))\exp(-\theta_i)\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))\left\{y_i\theta_i-\log(1+\exp(\theta_i))\right\}-$$

$$\left.(1+\exp(\theta_i))\exp(-\theta_i)\left\{y_i\theta_i-\log(1+\exp(\theta_i))\right\}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_0}=\sum_{i=1}^{n}\left\{(2+\exp(-\theta_i)+\exp(\theta_i))\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.\left\{y_i\theta_i-\log(1+\exp(\theta_i))\right\}(\exp(\theta_i)-\exp(-\theta_i))\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_0}=0$

$$\sum_{i=1}^{n}\left\{(2+\exp(-\theta_i)+\exp(\theta_i))\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.\left\{y_i\theta_i-\log(1+\exp(\theta_i))\right\}(\exp(\theta_i)-\exp(-\theta_i))\right\}=0 \qquad (6.23)$$

While $\theta_i=\beta_0+\beta_1x_{fi}+\beta_2x_{gi}$

Again differentiate (6.22) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_1}=\frac{\partial}{\partial\beta_1}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i-\log(1+\exp(\theta_i)\}]$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_ix_{fi}-\frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))x_{fi}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}-$$

$$\left.(1+\exp(\theta_i))\exp(-\theta_i)x_{fi}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}-$$

$$\left.(1+\exp(\theta_i))\exp(-\theta_i)x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\{\exp(\theta_i)(1+\exp(-\theta_i))-(1+\exp(\theta_i))\exp(-\theta_i)\}$$

$$\left.x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}.$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_1}=\sum_{i=1}^{n}\left\{(2+\exp(-\theta_i)+\exp(\theta_i))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}(\exp(\theta_i)-\exp(-\theta_i))\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_1}=0$

$$\sum_{i=1}^{n}\left\{(2+\exp(-\theta_i)+\exp(\theta_i))\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}(\exp(\theta_i)-\exp(-\theta_i))\right\}=0 \tag{6.24}$$

Again differentiate (6.22) with respect to β_2 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=\frac{\partial}{\partial\beta_2}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i-\log(1+\exp(\theta_i)\}]$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_i x_{gi}-\frac{x_{gi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))x_{gi}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}-$$

$$(1+\exp(\theta_i))\exp(-\theta_i)x_{gi}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}\Bigg\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_ix_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\exp(\theta_i)(1+\exp(-\theta_i))x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}-$$

$$\left.(1+\exp(\theta_i))\exp(-\theta_i)x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}$$

$$=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_ix_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\{\exp(\theta_i)(1+\exp(-\theta_i))-(1+\exp(\theta_i))\exp(-\theta_i)\}$$

$$\left.x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_ix_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}.(\exp(\theta_i)-\exp(-\theta_i))\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=0$

$$\sum_{i=1}^{n}\left\{(1+\exp(-\theta_i))(1+\exp(\theta_i))\left\{y_ix_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}(\exp(\theta_i)-\exp(-\theta_i))\right\}=0 \qquad (6.25)$$

while $\theta_i=\beta_0+\beta_1x_{fi}+\beta_2x_{gi}$

now the numerical solution of the above equations (6.23), (6.24) and (6.25) provide us the posterior modes of β_0, β_1 and β_2.

6.3.3.3 Joint Posterior Distribution Using Jeffreys Prior

Now for the joint posterior distribution of β_0, β_1 and β_2 we consider the joint Log Likelihood function (6.7).and the joint Jeffreys prior (6.11), then the joint posterior distribution of β_0, β_1 and β_2 is found to be:

$$p(\beta_0,\beta_1,\beta_2\,|\,data) \propto \sum_{i=1}^{n}\{y_i\log(p_i)+(1-y_i)\log(1-p_i)\}p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}}$$

$$p(\beta_0,\beta_1,\beta_2\,|\,data) \propto \sum_{i=1}^{n}\left\{y_i\log\left(\frac{p_i}{1-p_i}\right)+\log(1-p_i)\right\}p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}} \qquad (6.26)$$

By using logistic regression model with two explanatory given in equation (6.1) see section (6.3) and the expression derived from (6.1) that is equation (6.15):

Then after replacing equation (6.1) and (6.15) the above posterior distribution (6.26) becomes:

$$p(\beta_0,\beta_1,\beta_2\,|\,data) \propto \sum_{i=1}^{n}\{y_i(\beta_0+\beta_1x_{fi}+\beta_2x_{gi})-\log\{1+\exp(\beta_0+\beta_1x_{fi}+\beta_2x_{gi})\}\}p_i^{-\frac{1}{2}}(1-p_i)^{-\frac{1}{2}}$$

Now we will also use equation (6.1) to derive expressions given as:

$$p_i^{-\frac{1}{2}}=\sqrt{1+\exp(-(\beta_0+\beta_1x_{fi}+\beta_2x_{gi}))} \;\; \& \;\; (1-p_i)^{-\frac{1}{2}}=\sqrt{1+\exp(\beta_0+\beta_1x_{fi}+\beta_2x_{gi})} \qquad (6.27)$$

As we know we have already suppose that $\theta_i=\beta_0+\beta_1x_{fi}+\beta_2x_{gi}$ then the joint Posterior distribution for β_0, β_1 and β_2 will become as:

$$p(\beta_0,\beta_1,\beta_2\,|\,data) \propto \sum_{i=1}^{n}\{y_i\theta_i-\log\{1+\exp(\theta_i)\}\}\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}$$

$$p(\beta_0,\beta_1,\beta_2 \mid data)=\frac{1}{k}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\} \qquad (6.28)$$

$$-\infty<\beta_0<\infty,\ -\infty<\beta_1<\infty,-\infty<\beta_2<\infty$$

This is the joint posterior distribution of β_0, β_1 and β_2, where k is the normalizing constant. Now to estimate the unknown parameters partially differentiate the above equation (6.28) with respect to β_0, β_1 and β_2 then equating to zero. So this numerical solution will provide us the Posterior estimates (modes), for this we precede as follows:

Differentiate (6.28) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0}=\frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i-\frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i-\frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left\{\frac{\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}}{2\sqrt{1+\exp(\theta_i)}}-\frac{\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$\{y_i\theta_i - \log(1+\exp(\theta_i))\}\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = \sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i - \frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.\{y_i\theta_i - \log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i - \frac{1}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.\{y_i\theta_i - \log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\} = 0 \qquad (6.29)$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

Again differentiate (6.28) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right\}$$

$$= \sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i x_{fi} - \frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\frac{x_{fi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\left\{y_i\theta_i + \log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(\theta_i)}} -$$

$$\left.\frac{x_{fi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\frac{x_{fi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{x_{fi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_1}=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_1}=0$

$$\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{fi}-\frac{x_{fi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{fi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}=0 \qquad (6.30)$$

Again differentiate (6.28) with respect to β_2 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=\frac{\partial}{\partial\beta_2}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{1+\exp(-\theta_i)}\sqrt{1+\exp(\theta_i)}\left\{y_i x_{gi}-\frac{x_{gi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)}\right\}+\right.$$

$$\frac{x_{gi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{x_{gi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\left\{y_i\theta_i+\log\left(1-\frac{1}{1+\exp(-\theta_i)}\right)\right\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\frac{x_{gi}\exp(\theta_i)\sqrt{1+\exp(-\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(\theta_i)}}-$$

$$\left.\frac{x_{gi}\exp(-\theta_i)\sqrt{1+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}}{2\sqrt{1+\exp(-\theta_i)}}\right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$\left.x_{gi}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2\mid data)}{\partial\beta_2}=0$

$$\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\left\{y_i x_{gi}-\frac{x_{gi}}{1+\exp(-\theta_i)}\right\}+\right.$$

$$x_{gi}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\left\{\frac{\exp(\theta_i)-\exp(-\theta_i)}{2\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}}\right\}\Bigg\} = 0 \qquad (6.31)$$

6.3.3.4 Joint Posterior distribution Using Uniform Prior

Now using the joint Log likelihood function (6.7).and the joint uniform prior distribution (6.13), then the joint posterior distribution of β_0, β_1 and β_2 is found to be:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i\log(p_i)+(1-y_i)\log(1-p_i)\}1$$

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\left\{y_i\log\left(\frac{p_i}{1-p_i}\right)+\log(1-p_i)\right\} \qquad (6.32)$$

By using logistic regression model with two explanatory given in equation (6.1) see section (6.3) and the expression derived from (6.1) that is equation (6.15):

Then after replacing equation (6.1) and (6.15) the above posterior distribution (6.32) becomes:

$$p(\beta_0,\beta_1,\beta_2 \mid data) \propto \sum_{i=1}^{n}\{y_i(\beta_0+\beta_1x_{fi}+\beta_2x_{gi})-\log\{1+\exp(\beta_0+\beta_1x_{fi}+\beta_2x_{gi})\}\}$$

As we know we have already suppose that $\theta_i = \beta_0+\beta_1x_{fi}+\beta_2x_{gi}$ then the joint Posterior distribution for β_0, β_1 and β_2 will become as:

$$p(\beta_0,\beta_1,\beta_2 \mid data) = \frac{1}{k}\sum_{i=1}^{n}[y_i\theta_i - \log\{1+\exp(\theta_i)\}] \qquad (6.33)$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

This is the joint posterior distribution of β_0, β_1 and β_2, where k is the normalizing constant. Here our main objective is to estimate these unknown parameters. Then for this purpose if we partially differentiate the above equation (6.33) with respect to β_0, β_1 and β_2

simultaneously and equate it to zero. So this numerical solution will provide us the Posterior modes, so for this we proceed as follows.

Differentiate (6.33) with respect to β_0 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\sum_{i=1}^{n}\left[y_i\theta_i - \log\{1+\exp(\theta_i)\}\right]$$

$$= \sum_{i=1}^{n}\left\{ y_i - \frac{\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)} \right\}$$

$$= \sum_{i=1}^{n}\left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = \sum_{i=1}^{n}\left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_0} = 0$

$$\sum_{i=1}^{n}\left\{ y_i - \frac{1}{1+\exp(-\theta_i)} \right\} = 0 \qquad (6.34)$$

Now differentiate (6.33) with respect to β_1 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\left[y_i\theta_i - \log\{1+\exp(\theta_i)\}\right]$$

$$= \sum_{i=1}^{n}\left\{ y_i x_{fi} - \frac{x_{fi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)} \right\}$$

$$= \sum_{i=1}^{n}\left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_1} = \sum_{i=1}^{n}\left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_1} = 0$

$$\sum_{i=1}^{n}\left\{ y_i x_{fi} - \frac{x_{fi}}{1+\exp(-\theta_i)} \right\} = 0 \qquad (6.35)$$

Now again differentiate (6.33) with respect to β_2 we obtain:

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\sum_{i=1}^{n}\left[y_i\theta_i - \log\{1+\exp(\theta_i)\}\right]$$

$$= \sum_{i=1}^{n}\left\{ y_i x_{gi} - \frac{x_{gi}\exp(-\theta_i)}{(1+\exp(-\theta_i))^2\left(1-\frac{1}{1+\exp(-\theta_i)}\right)} \right\}$$

$$= \sum_{i=1}^{n}\left\{ y_i x_{gi} - \frac{x_{gi}}{1+\exp(-\theta_i)} \right\}$$

$$\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_2} = \sum_{i=1}^{n}\left\{ y_i x_{gi} - \frac{x_{gi}}{1+\exp(-\theta_i)} \right\}$$

Now for maximizing put $\frac{\partial p(\beta_0,\beta_1,\beta_2 \mid data)}{\partial \beta_2} = 0$

$$\sum_{i=1}^{n}\left\{ y_i x_{gi} - \frac{x_{gi}}{1+\exp(-\theta_i)} \right\} = 0 \qquad (6.36)$$

6.4 Selection of Hyperparameters

Since we know that the prior distributions of parameters β_0, β_1 & β_2 are as follows $\beta_0 \sim N(a_0, b_0)$, $\beta_1 \sim N(a_1, b_1)$ and $\beta_2 \sim N(a_2, b_2)$. Our main objective here is to find the values of these hyperparameters as a_0, a_1 & a_2 are the means of prior distributions and b_0, b_1 & b_2

are the variances of the prior distributions. We have suggested a range of values of hyperparameters and suggest the values with minimum standard error.

Table 6.2
Posterior Estimates at Different Values of Hyperparameters

Hyperparameters						**Posterior Mode**		
Mean a_0	*Var* b_0	*Mean* a_1	*Var* b_1	*Mean* a_2	*Var* b_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0	1	0	1	0	1	−12.5058	2.0321	0.1474
38	12.50	10.50	7.50	8.50	3.50	−30.4570	6.6342	1.7525
35	10.25	9.50	6.50	7.50	2.75	−26.9784	5.7213	1.2549
32.50	8.75	8.25	5.25	6.50	2.25	−24.2398	4.1282	1.0979
29.75	7.25	7.75	4.25	5.25	1.90	−21.8572	3.4575	0.9258
25.75	6.50	6.50	3.75	4.50	1.45	−16.2976	2.8974	0.7549
21.50	5.75	5.25	3.25	3.75	1.10	−12.9747	2.2974	0.5940
19.25	5.25	4.90	3.10	3.50	0.95	−10.5321	1.8878	0.3972
18.95	5.15	4.75	3.05	3.25	0.75	−9.9371	1.6378	0.1336

These are the values for hyperparameters for informative priors which are Normal priors for each parameter that is considered independent. Where *mean* = a_0 and *variance* = b_0 for the prior distribution of β_0 while *mean* = a_1 and *variance* = b_1 for the prior distribution of β_1 and mean = a_2 and variance = b_2 for the prior distribution of β_2. We suggest different values for the hyperparameters and find the values of posterior modes. So finally we decided to select the values of hyperparameters as *mean* = *18.95* and *variance* = *5.15* for the prior distribution of β_0, *mean* = *4.75* and *variance* = *3.05* for the prior distribution of β_1 and *mean* = *3.25* and *variance* = *0.75* for the prior distribution of β_2 and used these values for further Bayesian analysis.

6.5 Bayesian Analysis with Informative and Noninformative Priors

In this section we will present the Bayesian analysis with informative and noninformative priors. The analysis is based on the posterior distributions that are derived in previous sections:

6.5.1 Bayesian Analysis Using Joint Normal Prior

In this section we will present the Bayesian analysis of binary logistic regression model with two explanatory variables by using informative (Joint Normal) prior. Then the joint posterior distribution for the parameters $\beta_0, \beta_1 \& \beta_2$ derived in section (6.4.3.1) see equation (6.16):

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k} \sum_{i=1}^{n} \left[\exp\left(-\frac{\phi^*}{2b^*} \right) \{ y_i \theta_i - \log\{1 + \exp(\theta_i)\} \} \right]$$

$$-\infty < \beta_0 < \infty, \ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

where k is the normalizing constant:

Programs in SAS package have been designed similar program is given in appendix IV to show the graph of marginal densities of the parameters $\beta_0, \beta_1 \& \beta_2$ by using the data set given in Table 6.1.

Graph of Posterior Marginal Densities using Normal Prior

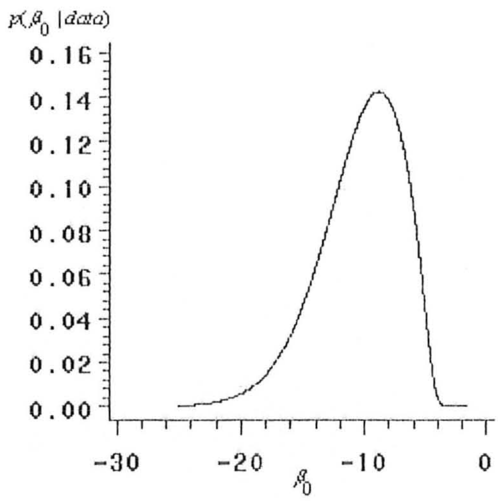

Figure 1(a)

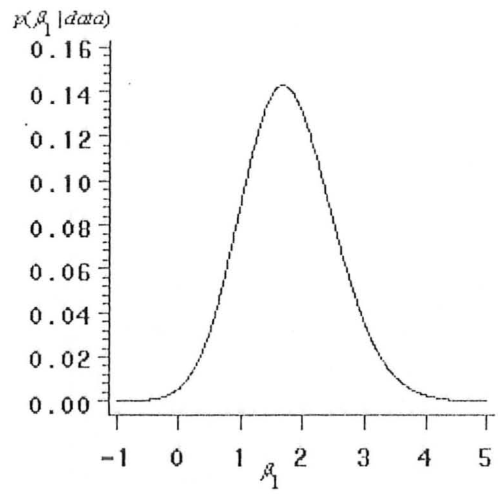

Figure 1(b)

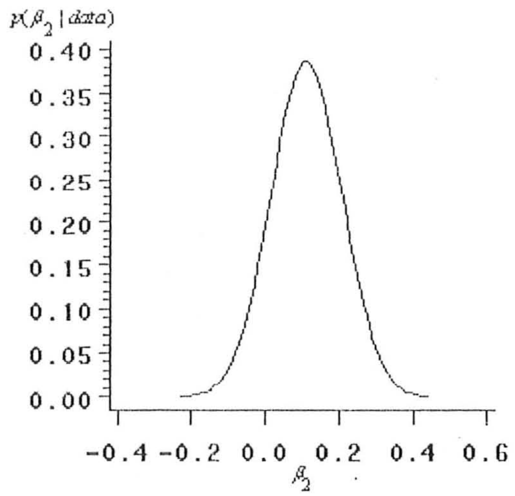

Figure 1(c)

6.5.1.1 Posterior Estimates

For further analysis we have designed a program in SAS package, program is given in appendix III and use the data set given in Table 6.1 and use the hyperparameters obtained in section (6.4.2). We have used Marquart method to obtain the posterior modes; while Quadrature method is used to obtain posterior means and standard errors.

Table 6.3

Posterior Estimates Using Joint Normal Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	−12.2557	−9.9371	5.1479		−0.4504
$\hat{\beta}_1$	2.0125	1.6378	0.8245	5.1438	0.4545
$\hat{\beta}_2$	0.1582	0.1336	0.1109	1.1429	0.2218

Here we have observed that the posterior mode for β_0 is greater then the posterior mean of β_0 which indicates that the distribution of this parameter is negatively skewed how much it is skewed we have calculated the coefficient of skewness given in Table 6.3. This can also be observed in figure 1(a). We have also observed that the posterior mean of β_1 is greater then the posterior mode of β_1 which shows that the distribution of parameter is positively skewed how much it is skewed we have calculated the coefficient of skewness, that is given in Table 6.3, and the graph in figure 1(b). We have also observed for β_2 that also shows a positively skewed but coefficient of skewness is much smaller than β_1, so it can said that the parameter β_2 has less skewed distribution than the distribution in parameter β_1. It is observe that the odds ratio is greater then 1 for both parameters which indicate that both the variables are more likely to occur, so the odds ratio is high for a healthy individual with ESR less than 20 mm\h to become an unhealthy or abnormal case with ESR greater than or equal to 20 mm/h. So it can be said that every one unit increase in the level of protein plasma (fibrinogen) approximately 5.1438 unit increases in the level of ESR and every one unit increase in the level of protein plasma (Y-globulin) approximately 1.1429 unit increase in the level of ESR, which is very low as compared to the other variable. So it can be concluded that the strength of relationship between the probability of an ESR reading greater than 20 mm/h and the level of protein plasma (fibrinogen) is high but for Y-globulin is very low. So it can be concluded that fibrinogen is very important variable to check any effect on the ESR as the level of this protein is rises but this is not the case with other variable that is Y-globulin, as its odds ratio is very low almost equal to one, which is not significant as far as the normal cases are concern.

6.5.1.2 Bayesian Hypothesis Testing

Hypotheses testing in Bayesian are very simple; here we only find the posterior probability by integrating the joint posterior distribution upon the parameters i.e.

We test the hypotheses:

$H_0 : \beta_1 \leq 0$ 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:

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

Now the posterior probability using informative prior while testing β_1 is:

$$p_0 = \int_{-5}^{0}\int_{-1}^{1}\int_{-24}^{24} \frac{1}{k}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right] d\beta_0 d\beta_2 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_0 = 0.001368$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.14% chance to accept H_0 so we accept H_1 with high probability and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma, this result provide the same conclusion as given with odds ratio.

The posterior probability for H_0' while testing β_2 is:

$$p_1 = p(\beta_2 \le 0) = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\beta_0, \beta_1, \beta_2 \mid data)\, d\beta_0 d\beta_1 d\beta_2$$

Now the posterior probability using informative prior while testing β_2 is:

$$p_1 = \int_{-1}^{0}\int_{-5}^{5}\int_{-24}^{24} \frac{1}{k}\sum_{i=1}^{n}\left[\exp\left(-\frac{\phi^*}{2b^*}\right)\{y_i\theta_i - \log\{1+\exp(\theta_i)\}\}\right] d\beta_0 d\beta_1 d\beta_2$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package to find the posterior probability and after being run the program we find the posterior probability as:

$p_1 = 0.030028$

The posterior probability indicates that under Bayesian hypothesis criterion there is 3% chance to accept H_0 and conclude that β_2 is positive but not as significant as β_1 is; so it can be said that Y-globulin is does not significantly affect the ESR as it rises in the blood plasma.

6.5.2 Bayesian Analysis Using Joint Haldane Prior

In this section we will present the Bayesian analysis of binary logistic regression model with two explanatory variables by using noninformative (Joint Haldane) prior. Then the joint posterior distribution for the parameters $\beta_0, \beta_1 \,\&\, \beta_2$ derived in section (6.4.3.2) see equation (6.22):

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i - \log(1+\exp(\theta_i)\}\right]$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

where k is the normalizing constant:

Program in SAS package has been designed; similar program is given in appendix IV to show the graph of marginal densities of the parameters $\beta_0, \beta_1 \,\&\, \beta_2$ by using the data set given in Table 6.1.

Graph of Posterior Marginal Densities using Haldane Prior

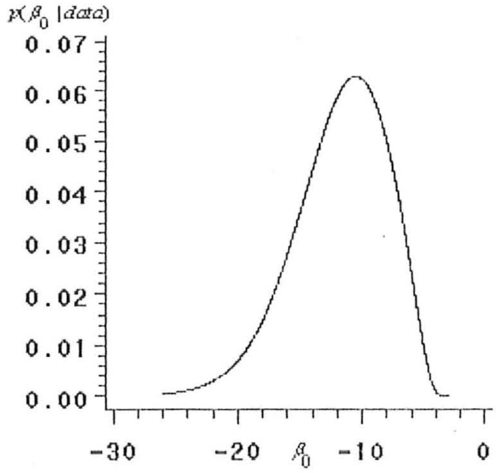

Figure 2(a)

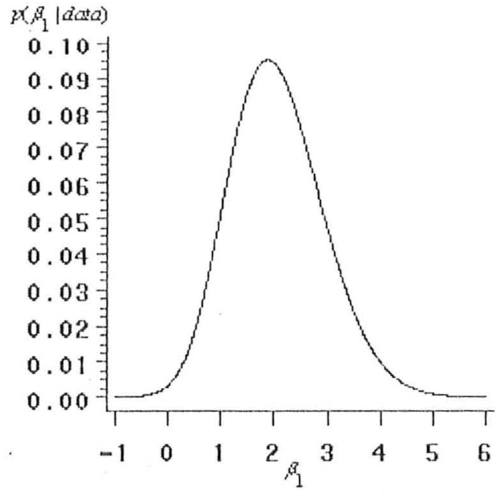

Figure 2(b)

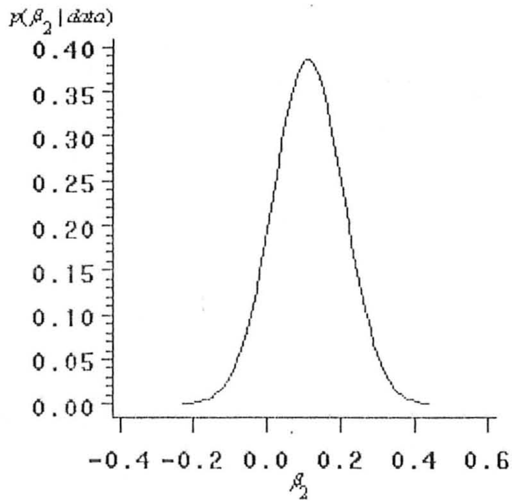

Figure 2 (c)

6.5.2.1 Posterior Estimates

For further analysis we have designed programs in SAS package, similar program is given in appendix III and also a similar program is given in appendix IV for standard error, while using the data set given in table 6.1. We have used Marquart method to obtain the posterior modes, while Quadrature method is used to obtain posterior means and standard errors by using the Posterior distribution:

Table 6.4

Posterior Estimates Using Joint Haldane Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	-12.8678	-10.4508	5.2374		-0.4615
$\hat{\beta}_1$	2.2641	1.8505	0.8759	6.3630	0.4722
$\hat{\beta}_2$	0.1625	0.1371	0.1116	1.1469	0.2276

We have similar results as we obtain in previous section β_0 has negatively skewed distribution how much it is skewed we have calculated the coefficient of skewness given in Table 6.4. Same can also be seen in figure 2(a). same is the case with parameter β_1 as its mean is greater than the posterior mode of β_1 which shows that the distribution of parameter is positively skewed how much it is skewed we have calculated the coefficient of skewness, that is given in Table 6.4, and figure 2(b) provide us with same results. We have also observed for β_2 that is also positively skewed but coefficient of skewness is much smaller than β_1, so it can said that the parameter β_2 has less skewed distribution than the distribution in parameter β_1. The odds ratio for this model is slightly different from the previous model but the significance is not changed. So that every one unit increase in the level of protein plasma (fibrinogen) approximately 6.3630 unit increases in the level of ESR and every one unit increase in the level of protein plasma (Y-globulin) approximately 1.1469 units increase in the level of ESR, which is very low. The conclusion is almost same as in previous section.

6.5.2.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterioi probability by integrating the joint posterior distribution upon the parameters i.e.

We test the hypothesis:

Now the posterior probability of the hypotheses given in section (6.5.1.2) using noninformative (Joint Haldane) prior while testing β_1 is:

$$p_0 = \int_{-6}^{0}\int_{-1}^{1}\int_{-26}^{26}\frac{1}{k}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i-\log(1+\exp(\theta_i)\}]d\beta_0 d\beta_2 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_0 = 0.001424$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.14% chance to accept H_0 so we accept H_1 with high probability and conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma:

Now the posterior probability of the hypotheses given in section (6.5.1.2) using noninformative (Haldane) prior is:

$$p_1 = \int_{-1}^{0}\int_{-6}^{6}\int_{-26}^{26}\frac{1}{k}\sum_{i=1}^{n}[\{2+\exp(-\theta_i)+\exp(\theta_i)\}\{y_i\theta_i-\log(1+\exp(\theta_i)\}]d\beta_0 d\beta_1 d\beta_2$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$p_1 = 0.030274$

The posterior probability indicates that under Bayesian hypothesis criterion there is 3% chance to accept H_0 and conclude that β_2 is positive but not as significant β_1 is; so it can be said that Y-globulin is does not significantly effect the ESR as its rise in the blood plasma:

6.5.3 Bayesian Analysis Using Joint Jeffreys Prior

In this section we will present the Bayesian analysis of binary logistic regression model with two explanatory variables by using noninformative (Joint Jeffreys) prior. Then the joint posterior distribution for the parameters β_0, β_1 & β_2 derived in section (6.4.3.3) see equation (6.28):

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right\}$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

Where k is the normalizing constant, Programs in SAS package have been designed; similar program is given in appendix IV to show the graph of marginal densities of the parameters β_0, β_1 & β_2 by using the data set given in Table 6.1.

Graph of Posterior Marginal Densities using Jeffreys Prior

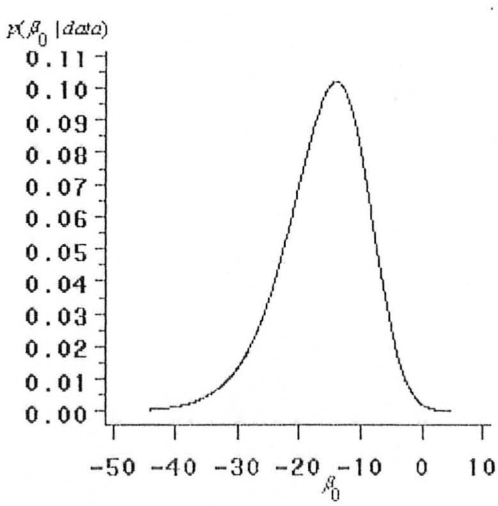

Figure 3(a)

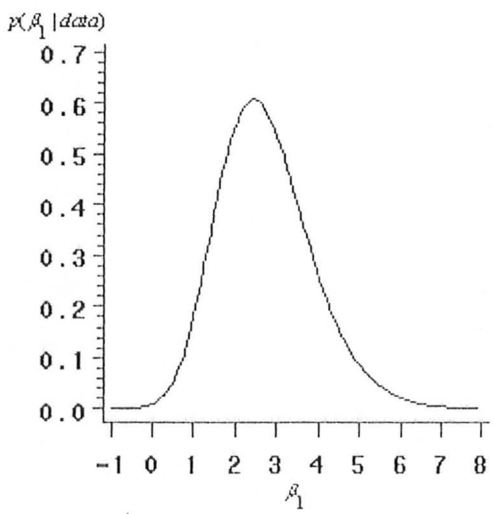

Figure 3(b)

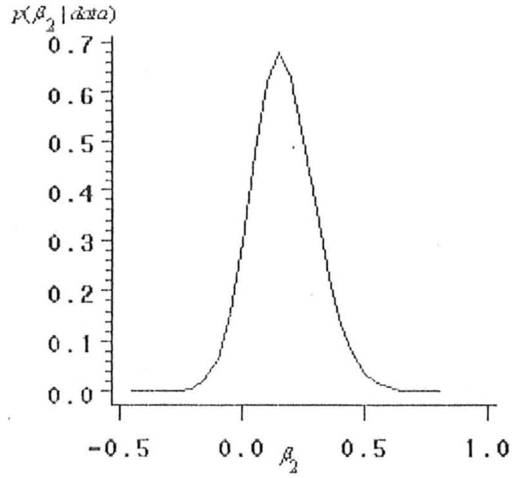

Figure 3(c)

6.5.3.1 Posterior Estimates

For further analysis we have designed program in SAS package, similar program is given in appendix III and use the data set given in table 6.1. We have used Marquart method to obtain the posterior modes; while Quadrature method is used to obtain posterior means and standard errors.

Table 6.5

Posterior Estimates Using Joint Jeffreys Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	-14.9810	-12.2886	5.5627		-0.4840
$\hat{\beta}_1$	2.4675	1.9506	0.9657	7.0329	0.5352
$\hat{\beta}_2$	0.1698	0.1428	0.1135	1.1535	0.2378

We have very similar results as we obtain in previous sections; β_0 has negatively skewed distribution how much it is skewed we have calculated the coefficient of skewness given in Table 6.4. Same is shown in figure 3(a). Same is the case with parameter β_1 as its mean is greater than the posterior mode of β_1 which shows that the distribution of parameter is positively skewed how much it is skewed we have calculated the coefficient of skewness, that is given in Table 6.4, and the graph in figure 3(b) provide us with same results. We have also observed for β_2 that also shows a positively skewed see Figure 3(c) but coefficient of skewness is much smaller than β_1, so it can said that the parameter β_2 has less skewed distribution than the distribution in parameter β_1. The odds ratio for this model is slightly different from the previous model but the significance is not changed. So that every one unit increase in the level of protein plasma (fibrinogen) approximately 7.0329 unit increases in the level of ESR and every one unit increase in the level of protein plasma (Y-globulin) approximately 1.1535 units increase in the level of ESR, which is very low. The conclusion is almost same as in previous section with a slight difference is results.

6.5.3.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the joint posterior distribution upon the parameters i.e.

The posterior probability of the hypotheses given in section (6.5.1.2) using noninformative (Jeffreys) prior while testing β_1 is:

$$p_0 = \int_{-8}^{0}\int_{-1}^{1}\int_{-41}^{-41}\frac{1}{k}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i - \log(1+\exp(\theta_i))\}\right\}d\beta_0 d\beta_2 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program has been designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.001473$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.15% chance to accept H_0 so accept H_1 with high probability and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma:

Now the posterior probability of hypotheses given in section (6.5.1.2) using noninformative (Jeffreys) prior while testing β_2 is:

$$p_1 = \int_{-1}^{0}\int_{-8}^{8}\int_{-41}^{41}\frac{1}{k}\sum_{i=1}^{n}\left\{\sqrt{2+\exp(-\theta_i)+\exp(\theta_i)}\{y_i\theta_i-\log(1+\exp(\theta_i))\}\right\}d\beta_0 d\beta_1 d\beta_2$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_1 = 0.031249$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 3% chance to accept H_0 and we conclude that β_2 is positive but not as significant β_1 is; so it can be said that Y-globulin is does not significantly effect the ESR as its rise in the blood plasma:

6.5.4 Bayesian Analysis Using Joint Uniform Prior

In this section we will present the Bayesian analysis of binary logistic regression model with two explanatory variables by using noninformative (Joint Uniform) prior. Then the joint posterior distribution for the parameters β_0, β_1 & β_2 derived in section (6.4.3.4) see equation (6.33):

$$p(\beta_0, \beta_1, \beta_2 \mid data) = \frac{1}{k}\sum_{i=1}^{n}\{y_i\theta_i - \log(1+\exp(\theta_i))\}$$

$$-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$$

Where k is the normalizing constant:

Programs in SAS package has been designed to show the graph of marginal densities of the parameters β_0, β_1 & β_2 by using the data set given in Table 6.1.

Graph of Posterior Marginal Densities using Uniform Prior

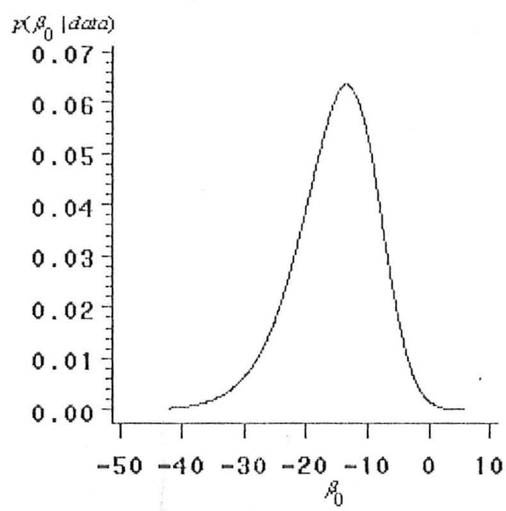

Figure 4(a)

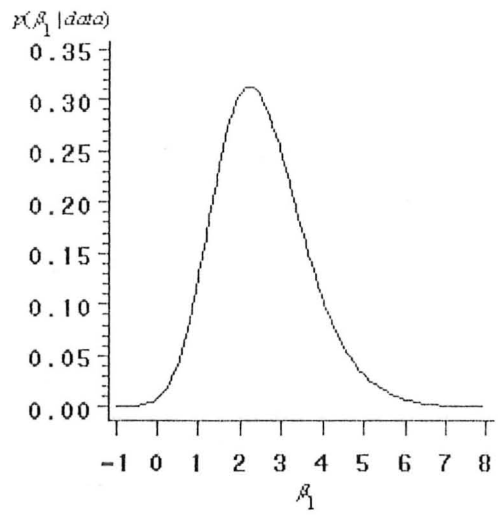

Figure 4(b)

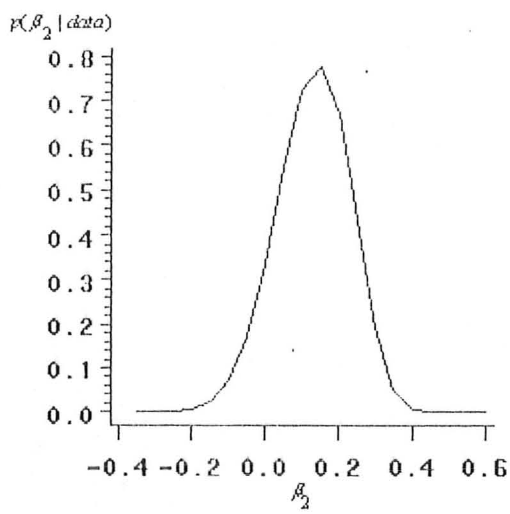

Figure 4(c)

6.5.4.1 Posterior Estimates

For further analysis we have designed a program in SAS package, similar program is given in appendix IV and use the data set given in table 6.1. We have used Marquart method to obtain the posterior modes; while Quadrature method is used to obtain posterior means and standard errors:

Table 6.6

Posterior Estimates Using Joint Uniform Prior

Regression Estimate	Posterior Mean	Posterior Mode	Standard Error	Odds Ratio	SK_p
$\hat{\beta}_0$	-15.2452	-12.5060	5.6480		-0.4850
$\hat{\beta}_1$	2.5658	2.0363	0.9811	7.6622	0.5397
$\hat{\beta}_2$	0.1727	0.1452	0.1151	1.1563	0.2389

We have slightly different results as we obtained in previous section, but with same conclusions, as β_0 has negatively skewed distribution how much it is skewed we have calculated the coefficient of skewness given in Table 6.4. Same is showing figure 4(a). same is the case with parameter β_1 as its mean is greater than the posterior mode of β_1 which shows that the distribution of parameter is positively skewed how much it is skewed we have calculated the coefficient of skewness, that is given in Table 6.4, and also figure 4(b) provide us with same results. We have also observed for β_2 that also shows a positively skewed but coefficient of skewness is much smaller than β_1, so it can said that the parameter β_2 has less skewed distribution than the distribution in parameter β_1, this can be seen in figure 4(c). The odds ratio for this model is slightly different from the previous model but the significance is not changed. So that every one unit increase in the level of protein plasma (fibrinogen) approximately 7.6622 unit increases in the level of ESR and every one unit increase in the level of protein plasma (Y-globulin) approximately 1.1563 units increase in the level of ESR, which is very low. The conclusion is almost same as in previous section.

6.5.4.2 Bayesian Hypothesis Testing

Hypothesis testing in Bayesian is very simple; here we only find the posterior probability by integrating the joint posterior distribution upon the parameters i.e.

We test the hypothesis:

The posterior probability of hypotheses given in section (6.5.1.2) using noninformative (Uniform) prior while testing β_1 is:

$$p_0 = \int_{-7}^{0}\int_{-1}^{1}\int_{-40}^{40} \frac{1}{k}\sum_{i=1}^{n}\{y_i\theta_i - \log(1+\exp(\theta_i))\}d\beta_0 d\beta_2 d\beta_1$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_0 = 0.001543$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 0.15% chance to accept H_0 and we conclude that β_1 is positive and playing a significant role to effect the ESR if this protein is rise in the blood plasma:

Now the posterior probability of hypotheses using informative prior while testing β_2 is:

$$p_1 = \int_{-1}^{0}\int_{-7}^{7}\int_{-40}^{40} \frac{1}{k}\sum_{i=1}^{n}\{y_i\theta_i - \log(1+\exp(\theta_i))\}d\beta_0 d\beta_1 d\beta_2$$

While $\theta_i = \beta_0 + \beta_1 x_{fi} + \beta_2 x_{gi}$

A program is designed in SAS package, similar program is given in appendix IV to find the posterior probability and after being run the program we find the posterior probability as:

$$p_1 = 0.031569$$

The posterior probability indicates that under Bayesian hypothesis criterion there is 3% chance to accept H_0 and we conclude that β_2 is positive but not as significant β_1 is; so it can be said that Y-globulin is does not significantly effect the ESR as its rise in the blood plasma:

6.6 Classical Regression Analysis

For the comparison purpose now we take the classical estimates and test the hypothesis. For this we have simply run the logistic regression without intercept model. Now the classical estimates and hypothesis testing is given in following section:

6.6.1 Classical Estimate

Using the data given in Table 6.1 and having run the logistic regression we obtain:

Table 6.7

Output of Logistic Regression Using Classical Approach

Coefficient	Classical Estimate	Standard Error	Z-Statistic	P-Value	Odds Ratio
$\hat{\beta}_0$	−12.5060	5.6480	−2.2100	0.0270	
$\hat{\beta}_1$	2.0323	0.9811	2.0800	0.0380	7.6600
$\hat{\beta}_2$	0.1452	0.1151	1.2600	0.2070	1.1600

6.6.2 Classical way of Hypotheses Testing

We have the logistic regression model as:

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

Hypotheses

$H_0 : \beta_1 \leq 0$ Versus $H_1 : \beta_1 > 0$

and

$H_0' : \beta_2 \leq 0$ Versus $H_1' : \beta_2 > 0$

Since the p-value for β_1 is 0.0380, it indicates that we accept H_0 at 3.8% level of significance. So it can be concluded that fibrinogen is playing significant role at 5% level of significance and it effect the ESR if this protein rises in the blood plasma. While the p-value for β_2 is 0.2070, it indicates that Y-globulin is not a significant variable at any level of significance, as its p-value is too large to support in favor of H_0' so it can be said that it does not effect ESR significantly if this protein rises in the blood plasma.

6.7 Comparison of Bayesian and Classical Logistic Regression Analysis

Now as a summery we present the results of logistic regression model with two explanatory variables we have obtained by using Bayesian and Classical techniques and make comparison between these two, the results are presented in table (6.8) are the Bayesian results obtained by using different priors these results can be compared with the results given in table (6.7).

Table 6.8

Posterior Estimates for Logistic Regression Model With Two Explanatory Variables

Coefficient		Noninformative Prior			Informative prior
		Uniform prior	Jeffreys prior	Haldane prior	
$\hat{\beta}_0$	Posterior Mode	−12.5060	−12.2886	−10.5408	−9.9371
	Posterior Mean	−15.2452	−14.9810	−12.8678	−12.2557
	Standard Error	5.6480	5.5627	5.2374	5.1479
	SK_p	−0.4850	−0.4840	−0.4615	−0.4504
$\hat{\beta}_1$	Posterior Mode	2.0363	1.9506	1.8505	1.6378
	Posterior Mean	2.5658	2.4675	2.2641	2.0125
	Odds Ratio	7.6622	7.0329	6.3630	5.1438
	Standard Error	0.9811	0.9657	0.8759	0.8245
	SK_p	0.5397	0.5352	0.4722	0.4545
$\hat{\beta}_2$	Posterior Mode	0.1452	0.1428	0.1371	0.1336
	Posterior Mean	0.1727	0.1698	0.1625	0.1582
	Odds Ratio	1.1563	1.1535	1.1469	1.1429
	Standard Error	0.1151	0.1135	0.1116	0.1109
	SK_p	0.2389	0.2378	0.2276	0.2218

The results found by using Classical logistic regression and in Bayesian logistic regression with Uniform prior are approximately same in all respects i.e. the coefficients, p-values and odds ratio. Here odds ratio are interpreted as the approximated change in the risk of disease for every one unit increase in the amount of fibrinogen and Y-globulin. So the results are much improved with Haldane and informative prior as compared to uniform and Jeffreys. At the end it can be said that Haldane prior performs better than Jeffrey's priors in binary logistic regression models, when we have skewed data sets, as the case for this particular data sets of ESR. While the best model we may suggest is the binary logistic regression model with intercept having only one explanatory variable that is fibrinogen which indicate a significant effect on ESR in all type logistic model used in our research as Y-globulin does not show any significant effect on ESR with informative and noninformative priors.

Chapter 7

Conclusion and Further Research

The present study comprises the Bayesian analysis of the binary logistic regression model taken from Cengiz et al. (2001). In Cengiz et al. (2001) linear regression model with two explanatory variables is given and they use approximation approach for the completion of their study and only consider noninformative priors for posterior analysis, but we have consider the model without intercept, with intercept and logistic regression model with two explanatory variables. We have considered the entire posterior distribution for Bayesian analysis i.e. no approximation is used. We have also obtained results by using informative prior. We have presented the Bayesian analysis of binary logistic regression model in different style. This analysis has been done using three noninformative (Uniform, Jeffreys and Haldane) priors and an informative (Normal) prior. The derivation of Haldane and Jeffreys prior is also provided. We have considered Bayesian testing of hypotheses about the parameters. The posterior probabilities for the hypotheses concerning to the parameters have been calculated. Then the decisions have been made about the hypotheses according to these posterior probabilities.

For informative prior, the hyperparameters are selected on the basis of expert opinion idea taken from Bian (1997). So for this purpose a range of values of hyperparameters are decided and selected the hyperparameters with minimum standard errors. The behavior of posterior estimates (modes & means) are also observed through graphical representation of marginal posteriors of parameters which shows some how a skewed pattern with different noninformative and informative priors.

The posterior means, the posterior modes, the standard errors, the odds ratio and the coefficient of skewness are obtained by designing the program in SAS package. These results are computed using data set provided by Cheng et al. (2001). The data set is used for Bayesian analysis of the models using Uniform, Jeffreys, Haldane and informative (Normal) priors. The results are also obtained by using Classical approach and are compared with the results obtained by Bayesian approach. We observe that the results that are obtained by Bayesian approach are more or less similar to that obtained by Classical approach. It is also observed that the results by using Uniform and Jeffreys prior are close to the results obtained by Classical approach but the results obtained by using Haldane and informative priors are slightly different. It is also observed that the variable Y-globulin is not effecting significantly to response variable (ESR). So the suggested model for further study is the model with intercept and with one explanatory variable that is Fibrinogen.

We have also presented the Bayesian hypotheses testing of binary logistic regression model for the data set given in Table 4.1 and Table 6.1 using informative (Normal) and noninformative (Uniform, Jeffreys and Haldane) priors and we have observe that the posterior probabilities suggest the same conclusion as provided by Classical results but in more significant way.

For further research, this work can be extended to many directions for the Bayesian analysis of logistic regression model. One may increase the number of independent variables; one may use Probit or Tobit method to handle these types of models that will provide same results as with logistic with slight differences. One may also consider the ordered or multinomial categories for response variable by using ordered Logit, ordered Probit or ordered Tobit model to handle this and multinomial Logit, multinomial Probit or multinomial Tobit model to handle multinomial responses. This work can also be extended for different informative priors that may be Conjugate. Different techniques for the elicitation of hyperparameters can also be used for informative prior.

Appendix I

```
*TO FIND THE POSTERIOR MODE USING INFORMATIVE PRIOR WITHOUT
INTERCEPT;

DATA DD;
INPUT NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
Y19 Y20 Y21 Y22 Y23  Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131
X132 a1 b1 C1;
CARDS;
32 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
2.52 2.56 2.19 2.18 3.41 2.46 3.22 2.21 3.15 2.60 2.29 2.35 5.06 3.34 2.38 3.15 3.53
2.68 2.60 2.23 2.88 2.65 2.09 2.28 2.67 2.29 2.15 2.54 3.93 3.34 2.99 3.32 1 0.20 0
;
PROC PRINT DATA=DD; RUN;
PROC SYSNLIN DATA=DD;
NP11=EXP(-B1*X11); NP12=EXP(-B1*X12); NP13=EXP(-B1*X13);
NP14=EXP(-B1*X14); NP15=EXP(-B1*X15); NP16=EXP(-B1*X16);
NP17=EXP(-B1*X17); NP18=EXP(-B1*X18); NP19=EXP(-B1*X19);
NP110=EXP(-B1*X110); NP111=EXP(-B1*X111); NP112=EXP(-B1*X112);
NP113=EXP(-B1*X113); NP114=EXP(-B1*X114); NP115=EXP(-B1*X115);
NP116=EXP(-B1*X116); NP117=EXP(-B1*X117); NP118=EXP(-B1*X118);
NP119=EXP(-B1*X119); NP120=EXP(-B1*X120); NP121=EXP(-B1*X121);
NP122=EXP(-B1*X122); NP123=EXP(-B1*X123); NP124=EXP(-B1*X124);
NP125=EXP(-B1*X125); NP126=EXP(-B1*X126); NP127=EXP(-B1*X127);
NP128=EXP(-B1*X128); NP129=EXP(-B1*X129); NP130=EXP(-B1*X130);
NP131=EXP(-B1*X131); NP132=EXP(-B1*X132); PP11=EXP(B1*X11);
PP12=EXP(B1*X12); PP13=EXP(B1*X13); PP14=EXP(B1*X14);
PP15=EXP(B1*X15); PP16=EXP(B1*X16); PP17=EXP(B1*X17);
PP18=EXP(B1*X18); PP19=EXP(B1*X19); PP110=EXP(B1*X110);
PP111=EXP(B1*X111); PP112=EXP(B1*X112); PP113=EXP(B1*X113);
PP114=EXP(B1*X114); PP115=EXP(B1*X115); PP116=EXP(B1*X116);
PP117=EXP(B1*X117); PP118=EXP(B1*X118); PP119=EXP(B1*X119);
PP120=EXP(B1*X120); PP121=EXP(B1*X121); PP122=EXP(B1*X122);
PP123=EXP(B1*X123); PP124=EXP(B1*X124); PP125=EXP(B1*X125);
PP126=EXP(B1*X126); PP127=EXP(B1*X127); PP128=EXP(B1*X128);
PP129=EXP(B1*X129); PP130=EXP(B1*X130); PP131=EXP(B1*X131);
PP132=EXP(B1*X132); K=EXP(-1/2*1/b1**2*(B1-a1)**2);
C1=(K*(X11*Y1-1/1+NP11)-(B1-a1)/b1**2*(Y1*(B1*X11))+LOG(1/1+PP11)+
   (X12*Y2-1/1+NP12)-(B1-a1)/b1**2*(Y2*(B1*X12))+LOG(1/1+PP12)+
   (X13*Y3-1/1+NP13)-(B1-a1)/b1**2*(Y3*(B1*X13))+LOG(1/1+PP13)+
   (X14*Y4-1/1+NP14)-(B1-a1)/b1**2*(Y4*(B1*X14))+LOG(1/1+PP14)+
   (X15*Y5-1/1+NP15)-(B1-a1)/b1**2*(Y5*(B1*X15))+LOG(1/1+PP15)+
   (X16*Y6-1/1+NP16)-(B1-a1)/b1**2*(Y6*(B1*X16))+LOG(1/1+PP16)+
```

```
  (X17*Y7-1/1+NP17)-(B1-a1)/b1**2*(Y7*(B1*X17))+LOG(1/1+PP17)+
  (X18*Y8-1/1+NP18)-(B1-a1)/b1**2*(Y8*(B1*X18))+LOG(1/1+PP18)+
  (X19*Y9-1/1+NP19)-(B1-a1)/b1**2*(Y9*(B1*X19))+LOG(1/1+PP19)+
  (X110*Y10-1/1+NP110)-(B1-a1)/b1**2*(Y10*(B1*X110))+LOG(1/1+PP110)+
  (X111*Y11-1/1+NP111)-(B1-a1)/b1**2*(Y11*(B1*X111))+LOG(1/1+PP111)+
  (X112*Y12-1/1+NP112)-(B1-a1)/b1**2*(Y12*(B1*X112))+LOG(1/1+PP112)+
  (X113*Y13-1/1+NP113)-(B1-a1)/b1**2*(Y13*(B1*X113))+LOG(1/1+PP113)+
  (X114*Y14-1/1+NP114)-(B1-a1)/b1**2*(Y14*(B1*X114))+LOG(1/1+PP114)+
  (X115*Y15-1/1+NP115)-(B1-a1)/b1**2*(Y15*(B1*X115))+LOG(1/1+PP115)+
  (X116*Y16-1/1+NP116)-(B1-a1)/b1**2*(Y16*(B1*X116)+LOG(1/1+PP116)+
  (X117*Y17-1/1+NP117)-(B1-a1)/b1**2*(Y17*(B1*X117))+LOG(1/1+PP117)+
  (X118*Y18-1/1+NP118)-(B1-a1)/b1**2*(Y18*(B1*X118))+LOG(1/1+PP118)+
  (X119*Y19-1/1+NP119)-(B1-a1)/b1**2*(Y19*(B1*X119))+LOG(1/1+PP119)+
  (X120*Y20-1/1+NP120)-(B1-a1)/b1**2*(Y20*(B1*X120))+LOG(1/1+PP120)+
  (X121*Y21-1/1+NP121)-(B1-a1)/b1**2*(Y21*(B1*X121))+LOG(1/1+PP121)+
  (X122*Y22-1/1+NP122)-(B1-a1)/b1**2*(Y22*(B1*X122))+LOG(1/1+PP122)+
  (X123*Y23-1/1+NP123)-(B1-a1)/b1**2*(Y23*(B1*X123))+LOG(1/1+PP123)+
  (X124*Y24-1/1+NP124)-(B1-a1)/b1**2*(Y24*(B1*X124))+LOG(1/1+PP124)+
  (X125*Y25-1/1+NP125)-(B1-a1)/b1**2*(Y25*(B1*X125))+LOG(1/1+PP125)+
  (X126*Y26-1/1+NP126)-(B1-a1)/b1**2*(Y26*(B1*X126))+LOG(1/1+PP126)+
  (X127*Y27-1/1+NP127)-(B1-a1)/b1**2*(Y27*(B1*X127))+LOG(1/1+PP127)+
  (X128*Y28-1/1+NP128)-(B1-a1)/b1**2*(Y28*(B1*X128))+LOG(1/1+PP128)+
  (X129*Y29-1/1+NP129)-(B1-a1)/b1**2*(Y29*(B1*X129))+LOG(1/1+PP129)+
  (X130*Y30-1/1+NP130)-(B1-a1)/b1**2*(Y30*(B1*X130))+LOG(1/1+PP130)+
  (X131*Y31-1/1+NP131)-(B1-a1)/b1**2*(Y31*(B1*X131))+LOG(1/1+PP131)+
  (X132*Y32-1/1+NP132)-(B1-a1)/b1**2*(Y32*(B1*X132))+LOG(1/1+PP132)));
ENDO C1;
EXO NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131
X132;
PARMS B1 1; RUN;
```

Appendix II

```
* TO FIND THE POSTERIOR MODES USING JEFFREYS PRIOR WITH
INTERCEPT MODEL;

DATA DD;
INPUT NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131 X132 C0 C1;
CARDS;
32 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
2.52 2.56 2.19 2.18 3.41 2.46 3.22 2.21 3.15 2.60 2.29 2.35 5.06 3.34 2.38 3.15 3.53
2.68 2.60 2.23 2.88 2.65 2.09 2.28 2.67 2.29 2.15 2.54 3.93 3.34 2.99 3.32 0 0
;
PROC PRINT DATA=DD; RUN;
PROC SYSNLIN DATA=DD;
NP11=EXP(-B0-B1*X11);NP12=EXP(-B0-B1*X12);NP13=EXP(-B0-B1*X13);
NP14=EXP(-B0-B1*X14);NP15=EXP(-B0-B1*X15);NP16=EXP(-B0-B1*X16);
NP17=EXP(-B0-B1*X17);NP18=EXP(-B0-B1*X18);NP19=EXP(-B0-B1*X19);
NP110=EXP(-B0-B1*X110);NP111=EXP(-B0-B1*X111);NP112=EXP(-B0-B1*X112);
NP113=EXP(-B0-B1*X113);NP114=EXP(-B0-B1*X114);NP115=EXP(-B0-B1*X115);
NP116=EXP(-B0-B1*X116);NP117=EXP(-B0-B1*X117);NP118=EXP(-B0-B1*X118);
NP119=EXP(-B0-B1*X119);NP120=EXP(-B0-B1*X120);NP121=EXP(-B0-B1*X121);
NP122=EXP(-B0-B1*X122);NP123=EXP(-B0-B1*X123);NP124=EXP(-B0-B1*X124);
NP125=EXP(-B0-B1*X125);NP126=EXP(-B0-B1*X126);NP127=EXP(-B0-B1*X127);
NP128=EXP(-B0-B1*X128);NP129=EXP(-B0-B1*X129);NP130=EXP(-B0-B1*X130);
NP131=EXP(-B0-B1*X131);NP132=EXP(-B0-B1*X132);PP11=EXP(B0+B1*X11);
PP12=EXP(B0+B1*X12);PP13=EXP(B0+B1*X13);PP14=EXP(B0+B1*X14);
PP15=EXP(B0+B1*X15);PP16=EXP(B0+B1*X16);PP17=EXP(B0+B1*X17);
PP18=EXP(B0+B1*X18);PP19=EXP(B0+B1*X19);PP110=EXP(B0+B1*X110);
PP111=EXP(B0+B1*X111);PP112=EXP(B0+B1*X112);PP113=EXP(B0+B1*X113);
PP114=EXP(B0+B1*X114);PP115=EXP(B0+B1*X115);PP116=EXP(B0+B1*X116);
PP117=EXP(B0+B1*X117);PP118=EXP(B0+B1*X118);PP119=EXP(B0+B1*X119);
PP120=EXP(B0+B1*X120);PP121=EXP(B0+B1*X121);PP122=EXP(B0+B1*X122);
PP123=EXP(B0+B1*X123);PP124=EXP(B0+B1*X124);PP125=EXP(B0+B1*X125);
PP126=EXP(B0+B1*X126);PP127=EXP(B0+B1*X127);PP128=EXP(B0+B1*X128);
PP129=EXP(B0+B1*X129);PP130=EXP(B0+B1*X130);PP131=EXP(B0+B1*X131);
PP132=EXP(B0+B1*X132);PPP11=SQRT(2+NP11+PP11);PPP12=SQRT(2+NP12+PP
12);PPP13=SQRT(2+NP13+PP13);PPP14=SQRT(2+NP14+PP14);PPP15=SQRT(2+NP
15+PP15);PPP16=SQRT(2+NP16+PP16);PPP17=SQRT(2+NP17+PP17);PPP18=SQRT
(2+NP18+PP18);PPP19=SQRT(2+NP19+PP19);PPP110=SQRT(2+NP110+PP110);
PPP111=SQRT(2+NP111+PP111);PPP112=SQRT(2+NP112+PP112);PPP113=SQRT(2
+NP113+PP113);PPP114=SQRT(2+NP114+PP114);PPP115=SQRT(2+NP115+PP115);
PPP116=SQRT(2+NP116+PP116);PPP117=SQRT(2+NP117+PP117);PPP118=SQRT(2
+NP118+PP118);PPP119=SQRT(2+NP119+PP119);PPP120=SQRT(2+NP120+PP120);
```

```
    PPP121=SQRT(2+NP121+PP121);PPP122=SQRT(2+NP122+PP122);PPP123=SQRT(2
    +NP123+PP123);PPP124=SQRT(2+NP124+PP124);PPP125=SQRT(2+NP125+PP125);
    PPP126=SQRT(2+NP126+PP126);PPP127=SQRT(2+NP127+PP127);PPP128=SQRT(2
    +NP128+PP128);PPP129=SQRT(2+NP129+PP129);PPP130=SQRT(2+NP130+PP130);
    PPP131=SQRT(2+NP131+PP131);PPP132=SQRT(2+NP132+PP132);
C0=((PPP11*(Y1-1/(1+NP11))+(Y1*(B0+B1*X11)+LOG(1/(1+PP11)))*(PP11-NP11)/2*
PPP11)+(PPP12*(Y2-1/(1+NP12))+(Y2*(B0+B1*X12)+LOG(1/(1+PP12)))*(PP12-
NP12)/2*PPP12)+(PPP13*(Y3-1/(1+NP13))+(Y3*(B0+B1*X13)+LOG(1/(1+PP13)))*
(PP13-NP13)/2*PPP13)+(PPP14*(Y4-1/(1+NP14))+(Y4*(B0+B1*X14)+LOG(1/(1+
PP14)))*(PP14-NP14)/2*PPP14)+(PPP15*(Y5-1/(1+NP15))+(Y5*(B0+B1*X15)+LOG
(1/(1+PP15)))*(PP15-NP15)/2*PPP15)+(PPP16*(Y6-1/(1+NP16))+(Y6*(B0+B1*X16)
+LOG(1/(1+PP16)))*(PP16-NP16)/2*PPP16)+(PPP17*(Y7-1/(1+NP17))+(Y7*(B0+
B1*X17)+LOG(1/(1+PP17)))*(PP17-NP17)/2*PPP17)+(PPP18*(Y8-1/(1+NP18))+(Y8*
(B0+B1*X18)+LOG(1/(1+PP18)))*(PP18-NP18)/2*PPP18)+(PPP19*(Y9-1/(1+NP19))
+(Y9*(B0+B1*X19)+LOG(1/(1+PP19)))*(PP19-NP19)/2*PPP19)+(PPP110*(Y10-1/(1+
NP110))+(Y10*(B0+B1*X110)+LOG(1/(1+PP110)))*(PP110NP110)/2*PPP110)+(PPP111
*(Y11-1/(1+NP111))+(Y11*(B0+B1*X111)+LOG(1/(1+PP111)))*(PP111-NP111)/2*
PPP111)+(PPP112*(Y12-1/(1+NP112))+(Y12*(B0+B1*X112)+LOG(1/(1+PP112)))*
(PP112-NP112)/2*PPP112)+(PPP113*(Y131/(1+NP113))+(Y13*(B0+B1*X113)+LOG(1/
(1+PP113)))*(PP113-NP113)/2*PPP113)+(PPP114*(Y14-1/(1+NP114))+(Y14*(B0+B1*
X114)+LOG(1/(1+PP114)))*(PP114-NP114)/2*PPP114)+(PPP115*(Y15-1/(1+NP115))+
(Y15*(B0+B1*X115)+LOG(1/(1+PP115)))*(PP115-NP115)/2*PPP115)+(PPP116*(Y16-
1/(1+NP116))+(Y16*(B0+B1*X116)+LOG(1/(1+PP116)))*(PP116-NP116)/2*PPP116)+
(PPP117*(Y17-1/(1+NP117))+(Y17*(B0+B1*X117)+LOG(1/(1+PP117)))*(PP117-
NP117)/2*PPP117)+(PPP118*(Y18-1/(1+NP118))+(Y18*(B0+B1*X118)+LOG(1/(1+
PP118)))*(PP118-NP118)/2*PPP118)+(PPP119*(Y19-1/(1+NP119))+(Y19*(B0+B1*
X119)+LOG(1/(1+PP119)))*(PP119-NP119)/2*PPP119)+(PPP120*(Y20-1/(1+NP120))+
(Y20*(B0+B1*X120)+LOG(1/(1+PP120)))*(PP120-NP120)/2*PPP120)+(PPP121*(Y21-
1/(1+NP121))+(Y21*(B0+B1*X121)+LOG(1/(1+PP121)))*(PP121-NP121)/2*PPP121)+
(PPP122*(Y22-1/(1+NP122))+(Y22*(B0+B1*X122)+LOG(1/(1+PP122)))*(PP122-
NP122)/2*PPP122)+(PPP123*(Y23-1/(1+NP123))+(Y23*(B0+B1*X123)+LOG(1/(1+
PP123)))*(PP123-NP123)/2*PPP123)+(PPP124*(Y24-1/(1+NP124))+(Y24*(B0+
B1*X124)+LOG(1/(1+PP124)))*(PP124-NP124)/2*PPP124)+(PPP125*(Y25-1/(1+
NP125))+(Y25*(B0+B1*X125)+LOG(1/(1+PP125)))*(PP125-NP125)/2*PPP125)+
(PPP126*(Y26-1/(1+NP126))+(Y26*(B0+B1*X126)+LOG(1/(1+PP126)))*(PP126-
NP126)/2*PPP126)+(PPP127*(Y27-1/(1+NP127))+(Y27*(B0+B1*X127)+LOG(1/(1
+PP127)))*(PP127-NP127)/2*PPP127)+(PPP128*(Y28-1/(1+NP128))+(Y28*(B0+
B1*X128)+LOG(1/(1+PP128)))*(PP128-NP128)/2*PPP128)+(PPP129*(Y29-1/(1
+NP129))+(Y29*(B0+B1*X129)+LOG(1/(1+PP129)))*(PP129-NP129)/2*PPP129)+
(PPP130*(Y30-1/(1+NP130))+(Y30*(B0+B1*X130)+LOG(1/(1+PP130)))*(PP130-
NP130)/2*PPP130)+(PPP131*(Y31-1/(1+NP131))+(Y31*(B0+B1*X131)+LOG(1/
(1+PP131)))*(PP131-NP131)/2*PPP131)+(PPP132*(Y32-1/(1+NP132))+(Y32*(B0+B1*
X132)+LOG(1/(1+PP132)))*(PP132-NP132)/2*PPP132));
C1=((PPP11*X11*(Y1-1/(1+NP11))+X11*(Y1*(B0+B1*X11)+LOG(1/(1+PP11)))*(PP11
-NP11)/2*PPP11)+(PPP12*X12*(Y2-1/(1+NP12))+X12*(Y2*(B0+B1*X12)+LOG(1/(1
+PP12)))*(PP12-NP12)/2*PPP12)+(PPP13*X13*(Y3-1/(1+NP13))+X13*(Y3*(B0+B1*
```

```
X13)+LOG(1/(1+PP13)))*(PP13-NP13)/2*PPP13)+(PPP14*X14*(Y4-1/(1+NP14))+X14
*(Y4*(B0+B1*X14)+LOG(1/(1+PP14)))*(PP14-NP14)/2*PPP14)+(PPP15*X15*(Y5-1/(
1+NP15))+X15*(Y5*(B0+B1*X15)+LOG(1/(1+PP15)))*(PP15-NP15)/2*PPP15)+(PPP16
*X16*(Y6-1/(1+NP16))+X16*(Y6*(B0+B1*X16)+LOG(1/(1+PP16)))*(PP16-NP16)/2*
PPP16)+(PPP17*X17*(Y7-1/(1+NP17))+X17*(Y7*(B0+B1*X17)+LOG(1/(1+PP17)))*
(PP17-NP17)/2*PPP17)+(PPP18*X18*(Y8-1/(1+NP18))+X18*(Y8*(B0+B1*X18)+LOG
(1/(1+PP18)))*(PP18-NP18)/2*PPP18)+(PPP19*X19*(Y9-1/(1+NP19))+X19*(Y9*(B0+
B1*X19)+LOG(1/(1+PP19)))*(PP19-NP19)/2*PPP19)+(PPP110*X110*(Y10-1/(1+
NP110))+X110*(Y10*(B0+B1*X110)+LOG(1/(1+PP110)))*(PP110-NP110)/2*PPP110)
+(PPP111*X111*(Y11-1/(1+NP111))+X111*(Y11*(B0+B1*X111)+ LOG(1/(1+PP111)))
*(PP111-NP111)/2*PPP111)+(PPP112*X112*(Y12-1/(1+NP112))+X112*(Y12*(B0+
B1*X112)+LOG(1/(1+PP112)))*(PP112-NP112)/2*PPP112)+(PPP113*X113*(Y13-
1/(1+NP113))+X113*(Y13*(B0+B1*X113)+LOG(1/(1+PP113)))*(PP113-NP113)/2
*PPP113)+(PPP114*X114*(Y14-1/(1+NP114))+X114*(Y14*(B0+B1*X114)+LOG
(1/(1+PP114)))*(PP114-NP114)/2*PPP114)+(PPP115*X115*(Y15-1/(1+NP115))+X115
*(Y15*(B0+B1*X115)+LOG(1/(1+PP115)))*(PP115-NP115)/2*PPP115)+(PPP116*
X116*(Y16-1/(1+NP116))+X116*(Y16*(B0+B1*X116)+LOG(1/(1+PP116)))*(PP116-
NP116)/2*PPP116)+(PPP117*X117*(Y17-1/(1+NP117))+X117*(Y17*(B0+B1*X117)
+LOG(1/(1+PP117)))*(PP117-NP117)/2*PPP117)+(PPP118*X118*(Y18-1/(1+NP118
))+X118*(Y18*(B0+B1*X118)+LOG(1/(1+PP118)))*(PP118-NP118)/2*PPP118)+
(PPP119*X119*(Y19-1/(1+NP119))+X119*(Y19*(B0+B1*X119)+ LOG(1/(1+
PP119)))*(PP119-NP119)/2*PPP119)+(PPP120*X120*(Y20-1/(1+NP120))+X120*
(Y20*(B0+B1*X120)+LOG(1/(1+PP120)))*(PP120-NP120)/2*PPP120)+(PPP121*X121*
(Y21-1/(1+NP121))+X121*(Y21*(B0+B1*X121)+LOG(1/(1+PP121)))*(PP121-NP121)
/2*PPP121)+(PPP122*X122*(Y22-1/(1+NP122))+ X122*(Y22*(B0+B1*X122)+LOG
(1/(1+PP122)))*(PP122-NP122)/2*PPP122)+(PPP123*X123*(Y23-1/(1+NP123))+X123*
(Y23*(B0+B1*X123)+LOG(1/(1+PP123)))*(PP123-NP123)/2*PPP123)+(PPP124*
X124*(Y24-1/(1+NP124))+X124*(Y24*(B0+B1*X124)+LOG(1/(1+PP124)))*(PP124-
NP124)/2*PPP124)+(PPP125*X125*(Y25-1/(1+NP125)) +X125*(Y25*(B0+B1*X125)+
LOG(1/(1+PP125)))*(PP125-NP125)/2*PPP125)+(PPP126*X126*(Y26-1/(1+NP126))+
X126*(Y26*(B0+B1*X126)+LOG(1/(1+PP126)))*(PP126-NP126)/2*PPP126)+(PPP127*
X127*(Y27-1/(1+NP127))+X127*(Y27*(B0+B1*X127)+LOG(1/(1+PP127)))*(PP127-
NP127)/2*PPP127)+(PPP128*X128*(Y28-1/(1+NP128)) +X128*(Y28*(B0+B1*X128)+
LOG(1/(1+PP128)))*(PP128-NP128)/2*PPP128)+(PPP129*X129*(Y29-1/(1+NP129))+
X129*(Y29*(B0+B1*X129)+LOG(1/(1+PP129)))*(PP129-NP129)/2*PPP129)+(PPP130*
X130*(Y30-1/(1+NP130))+X130*(Y30*(B0+B1*X130)+LOG(1/(1+PP130)))*(PP130-
NP130)/2*PPP130)+(PPP131*X131*(Y31-1/(1+NP131))+ X131*(Y31*(B0+B1*X131)+
LOG(1/(1+PP131)))*(PP131-NP131)/2*PPP131)+(PPP132*X132*(Y32-1/(1+NP132))
+X132*(Y32*(B0+B1*X132)+LOG(1/(1+PP132)))*(PP132-NP132)/2*PPP132));
    ENDO C0 C1;
    EXO NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
    Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
    X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
    X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131 X132;
    PARMS B0 -1 B1 1; RUN;
```

Appendix III

```
* TO FIND THE POSTERIOR MODES USING INFORMATIVE PRIOR OF
LOGISTIC REGRESSION MODEL WITH TWO EXPLANATORY VARIABLES;

DATA DD;
INPUT NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131 X132 X21
X22 X23 X24 X25 X26 X27 X28 X29 X210 X211 X212 X213 X214 X215 X216 X217
X218 X219 X220 X221 X222 X223 X224 X225 X226 X227 X228 X229 X230 X231
X232 a0 a1 a2 b0 b1 b2 C0 C1 C2;
CARDS;
32 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
2.52 2.56 2.19 2.18 3.41 2.46 3.22 2.21 3.15 2.60 2.29 2.35 5.06 3.34 2.38 3.15 3.53
2.68 2.60 2.23 2.88 2.65 2.09 2.28 2.67 2.29 2.15 2.54 3.93 3.34 2.99 3.32 38 31 33 31
37 36 38 37 39 41 36 39 37 32 37 36 46 34 38 37 30 46 44 36 39 31 31 28 32 30 36 35
18.95 4.75 3.25 5.15 3.05 0.75 0 0 0
;
PROC PRINT DATA=DD; RUN;
PROC SYSNLIN DATA=DD;
NP11=EXP(-B0-B1*X11-B2*X21);NP12=EXP(-B0-B1*X12-B2*X22);NP13=EXP(-
B0-B1*X13-B2*X23);NP14=EXP(-B0-B1*X14-B2*X24);NP15=EXP(-B0-B1*X15-
B2*X25);NP16=EXP(-B0-B1*X16-B2*X26);NP17=EXP(-B0-B1*X17-B2*X27);
NP18=EXP(-B0-B1*X18-B2*X28);NP19=EXP(-B0-B1*X19-B2*X29);NP110=EXP(-
B0-B1*X110-B2*X210);NP111=EXP(-B0-B1*X111-B2*X211);NP112=EXP(-B0-
B1*X112-B2*X212);NP113=EXP(-B0-B1*X113-B2*X213);NP114=EXP(-B0-
B1*X114-B2*X214);NP115=EXP(-B0-B1*X115-B2*X215);NP116=EXP(-B0-
B1*X116-B2*X216);NP117=EXP(-B0-B1*X117-B2*X217);NP118=EXP(-B0-
B1*X118-B2*X218);NP119=EXP(-B0-B1*X119-B2*X219);NP120=EXP(-B0-
B1*X120-B2*X220);NP121=EXP(-B0-B1*X121-B2*X221);NP122=EXP(-B0-
B1*X122-B2*X222);NP123=EXP(-B0-B1*X123-B2*X223);NP124=EXP(-B0-
B1*X124-B2*X224);NP125=EXP(-B0-B1*X125-B2*X225);NP126=EXP(-B0-
B1*X126-B2*X226);NP127=EXP(-B0-B1*X127-B2*X227);NP128=EXP(-B0-
B1*X128-B2*X228);NP129=EXP(-B0-B1*X129-B2*X229);NP130=EXP(-B0-
B1*X130-B2*X230);NP131=EXP(-B0-B1*X131-B2*X231);NP132=EXP(-B0-
B1*X132-B2*X232);PP11=EXP(B0+B1*X11+B2*X21);PP12=EXP(B0+B1
*X12+B2*X22);PP13=EXP(B0+B1*X13+B2*X23);PP14=EXP(B0+B1*X14+B2*X24)
;PP15=EXP(B0+B1*X15+B2*X25);PP16=EXP(B0+B1*X16+B2*X26);PP17=EXP(B0
+B1*X17+B2*X27);PP18=EXP(B0+B1*X18+B2*X28);PP19=EXP(B0+B1*X19+B2*
X29);PP110=EXP(B0+B1*X110+B2*X210);PP111=EXP(B0+B1*X111+B2*X211);PP
112=EXP(B0+B1*X112+B2*X212);PP113=EXP(B0+B1*X113+B2*X213);
PP114=EXP(B0+B1*X114+B2*X214);PP115=EXP(B0+B1*X115+B2*X215);PP116=
EXP(B0+B1*X116+B2*X216);PP117=EXP(B0+B1*X117+B2*X217);PP118=EXP(B0
+B1*X118+B2*X218);PP119=EXP(B0+B1*X119+B2*X219);PP120=EXP(B0+B1*X1
```

```
    20+B2*X220);PP121=EXP(B0+B1*X121+B2*X221);PP122=EXP(B0+B1*X122+B2*
    X222);PP123=EXP(B0+B1*X123+B2*X223);PP124=EXP(B0+B1*X124+B2*X224);P
    P125=EXP(B0+B1*X125+B2*X225);PP126=EXP(B0+B1*X126+B2*X226);
    PP127=EXP(B0+B1*X127+B2*X227);PP128=EXP(B0+B1*X128+B2*X228);PP129=
    EXP(B0+B1*X129+B2*X229);PP130=EXP(B0+B1*X130+B2*X230);PP131=EXP(B0
    +B1*X131+B2*X231);PP132=EXP(B0+B1*X132+B2*X232);
K=EXP(-1/2*1/b0**2*(B0-a0)**2-1/2*1/b1**2*(B1-a1)**2-1/2*1/b2**2*(B2-a2)**2);
C0=(K*(Y1-1/1+NP11)-(B0-a0)/b0**2*(Y1*(B0+B1*X11+B2*X21))+LOG(1/1+PP11)+
     (Y2-1/1+NP12)-(B0-a0)/b0**2*(Y2*(B0+B1*X12+B2*X22))+LOG(1/1+PP12)+
       (Y3-1/1+NP13)-(B0-a0)/b0**2*(Y3*(B0+B1*X13+B2*X23))+LOG(1/1+PP13)+
       (Y4-1/1+NP14)-(B0-a0)/b0**2*(Y4*(B0+B1*X14+B2*X24))+LOG(1/1+PP14)+
       (Y5-1/1+NP15)-(B0-a0)/b0**2*(Y5*(B0+B1*X15+B2*X25))+LOG(1/1+PP15)+
       (Y6-1/1+NP16)-(B0-a0)/b0**2*(Y6*(B0+B1*X16+B2*X26))+LOG(1/1+PP16)+
       (Y7-1/1+NP17)-(B0-a0)/b0**2*(Y7*(B0+B1*X17+B2*X27))+LOG(1/1+PP17)+
       (Y8-1/1+NP18)-(B0-a0)/b0**2*(Y8*(B0+B1*X18+B2*X28))+LOG(1/1+PP18)+
       (Y9-1/1+NP19)-(B0-a0)/b0**2*(Y9*(B0+B1*X19+B2*X29))+LOG(1/1+PP19)+
(Y10-1/1+NP110)-(B0-a0)/b0**2*(Y10*(B0+B1*X110+B2*X210))+LOG(1/1+PP110)+
(Y11-1/1+NP111)-(B0-a0)/b0**2*(Y11*(B0+B1*X111+B2*X211))+LOG(1/1+PP111)+
(Y12-1/1+NP112)-(B0-a0)/b0**2*(Y12*(B0+B1*X112+B2*X212))+LOG(1/1+PP112)+
(Y13-1/1+NP113)-(B0-a0)/b0**2*(Y13*(B0+B1*X113+B2*X213))+LOG(1/1+PP113)+
(Y14-1/1+NP114)-(B0-a0)/b0**2*(Y14*(B0+B1*X114+B2*X214))+LOG(1/1+PP114)+
(Y15-1/1+NP115)-(B0-a0)/b0**2*(Y15*(B0+B1*X115+B2*X215))+LOG(1/1+PP115)+
(Y16-1/1+NP116)-(B0-a0)/b0**2*(Y16*(B0+B1*X116+B2*X216))+LOG(1/1+PP116)+
(Y17-1/1+NP117)-(B0-a0)/b0**2*(Y17*(B0+B1*X117+B2*X217))+LOG(1/1+PP117)+
(Y18-1/1+NP118)-(B0-a0)/b0**2*(Y18*(B0+B1*X118+B2*X218))+LOG(1/1+PP118)+
(Y19-1/1+NP119)-(B0-a0)/b0**2*(Y19*(B0+B1*X119+B2*X219))+LOG(1/1+PP119)+
 (Y20-1/1+NP120)-(B0-a0)/b0**2*(Y20*(B0+B1*X120+B2*X220))+LOG(1/1+PP120)+
(Y21-1/1+NP121)-(B0-a0)/b0**2*(Y21*(B0+B1*X121+B2*X221))+LOG(1/1+PP121)+
(Y22-1/1+NP122)-(B0-a0)/b0**2*(Y22*(B0+B1*X122+B2*X222))+LOG(1/1+PP122)+
(Y23-1/1+NP123)-(B0-a0)/b0**2*(Y23*(B0+B1*X123+B2*X223))+LOG(1/1+PP123)+
(Y24-1/1+NP124)-(B0-a0)/b0**2*(Y24*(B0+B1*X124+B2*X224))+LOG(1/1+PP124)+
(Y25-1/1+NP125)-(B0-a0)/b0**2*(Y25*(B0+B1*X125+B2*X225))+LOG(1/1+PP125)+
(Y26-1/1+NP126)-(B0-a0)/b0**2*(Y26*(B0+B1*X126+B2*X226))+LOG(1/1+PP126)+
(Y27-1/1+NP127)-(B0-a0)/b0**2*(Y27*(B0+B1*X127+B2*X227))+LOG(1/1+PP127)+
(Y28-1/1+NP128)-(B0-a0)/b0**2*(Y28*(B0+B1*X128+B2*X228))+LOG(1/1+PP128)+
(Y29-1/1+NP129)-(B0-a0)/b0**2*(Y29*(B0+B1*X129+B2*X229))+LOG(1/1+PP129)+
(Y30-1/1+NP130)-(B0-a0)/b0**2*(Y30*(B0+B1*X130+B2*X230))+LOG(1/1+PP130)+
(Y31-1/1+NP131)-(B0-a0)/b0**2*(Y31*(B0+B1*X131+B2*X231))+LOG(1/1+PP131)+
(Y32-1/1+NP132)-(B0-a0)/b0**2*(Y32*(B0+B1*X132+B2*X232))+LOG(1/1+PP132));
C1=(K*(X11*Y1-1/1+NP11)-(B1-a1)/b1**2*(Y1*(B0+B1*X11+B2*X21))+
LOG(1/1+PP11)+(X12*Y2-1/1+NP12)-(B1-a1)/b1**2*(Y2*(B0+B1*X12+B2*X22))+
LOG(1/1+PP12)+(X13*Y3-1/1+NP13)-(B1-a1)/b1**2*(Y3*(B0+B1*X13+B2*X23))+
LOG(1/1+PP13)+(X14*Y4-1/1+NP14)-(B1-a1)/b1**2*(Y4*(B0+B1*X14+B2*X24))+
LOG(1/1+PP14)+(X15*Y5-1/1+NP15)-(B1-a1)/b1**2*(Y5*(B0+B1*X15+B2*X25))+
LOG(1/1+PP15)+(X16*Y6-1/1+NP16)-(B1-a1)/b1**2*(Y6*(B0+B1*X16+B2*X26))+
LOG(1/1+PP16)+(X17*Y7-1/1+NP17)-(B1-a1)/b1**2 *(Y7*(B0+B1*X17+B2*X27))+
```

LOG(1/1+PP17)+X18*Y8-1/1+NP18)-(B1-a1)/b1**2*(Y8*(B0+B1*X18+B2*X28))+
LOG(1/1+PP18)+ (X19*Y9-1/1+NP19)-(B1-a1)/b1**2*(Y9*(B0+B1*X19+B2*X29))+
LOG(1/1+PP19)+ (X110*Y10-1/1+NP110)-(B1-a1)/b1**2*(Y10*(B0+B1*X110+
B2*X210))+LOG(1/1+ PP110)+(X111*Y11-1/1+NP111)-(B1-a1)/b1**2*(Y11*
(B0+B1*X111+B2*X211))+LOG(1/1+PP111)+(X112*Y12-1/1+NP112)-(B1-a1)/b1**2
(Y12(B0+B1*X112+B2*X212))+LOG(1/1+PP112)+(X113*Y13-1/1+NP113)-(B1-
a1)/b1**2*(Y13*(B0+B1*X113 +B2*X213))+LOG(1/1+PP113)+(X114*Y14-1/1+NP114)-
(B1-a1)/b1**2*(Y14*(B0+B1*X114+B2*X214))+LOG(1/1+PP114)+(X115*Y15-1/1
+NP115)-(B1-a1)/b1**2*(Y15*(B0+B1*X115+B2*X215))+LOG(1/1+PP115)+
(X116*Y16-1/1+NP116)-(B1-a1)/b1**2 *(Y16*(B0+B1*X116+B2*X216))+
LOG(1/1+PP116)+(X117*Y17-1/1+NP117)-(B1-a1) /b1**2*(Y17*(B0+B1*X117+
B2*X217))+LOG(1/1+PP117)+(X118*Y18-1/1+NP118)-(B1-a1)/b1**2*(Y18*(B0+
B1*X118+B2*X218))+LOG(1/1+PP118)+(X119*Y19-1/1+ NP119)-(B1a1)/b1**2
(Y19(B0+B1*X119+B2*X219))+LOG(1/1+PP119)+(X120*Y20-1/1+NP120)-(B1-
a1)/b1**2*(Y20*(B0+B1*X120+B2*X220))+LOG(1/1+PP120)+(X121* Y21-1/1+NP121)-
(B1-a1)/b1**2*(Y21*(B0+B1*X121+B2*X221))+LOG(1/1+PP121)+
(X122*Y22-1/1+NP122)-(B1-a1)/b1**2*(Y22*(B0+B1*X122+B2*X222))+LOG(1/1+
PP122 +(X123*Y23-1/1+NP123)-(B1-a1)/b1**2*(Y23*(B0+B1*X123+B2*X223))+
LOG(1/1+PP123)+(X124*Y24-1/1+NP124)-(B1-a1)/b1**2*(Y24*(B0+B1*X124+B2*
X224))+LOG(1/1+PP124)+(X125*Y25-1/1+NP125)-(B1-a1)/b1**2*(Y25*(B0+B1*X125
+B2*X225))+LOG(1/1+PP125)+(X126*Y26-1/1+NP126)-(B1-a1)/b1**2*(Y26* (B0+B1*
X126+B2*X226))+LOG(1/1+PP126)+(X127*Y27-1/1+NP127)-(B1-a1)/b1**2*(Y27*
(B0+B1*X127+B2*X227))+LOG(1/1+PP127)+(X128*Y28-1/1+NP128)-(B1-a1)/b1**2
(Y28(B0+B1*X128+B2*X228))+LOG(1/1+PP128)+(X129*Y29-1/1+NP129)-(B1-
a1)/b1**2*(Y29*(B0+B1*X129+B2*X229))+LOG(1/1+PP129)+(X130*Y30-1/1+ NP130
)-(B1-a1)/b1**2*(Y30*(B0+B1*X130+B2*X230))+LOG(1/1+PP130)+(X131*Y31-1/1+
NP131)-(B1-a1)/b**2*(Y31*(B0+B1*X131+B2*X231))+LOG(1/1+PP131)+(X132*Y32-
1/1+NP132)-(B1-a1) /b1**2*(Y32*(B0+B1*X132+B2*X232))+LOG(1/1+PP132));
C2=(K*(X21*Y1-1/1+NP11)-(B2-a2)/b2**2*(Y1*(B0+B1*X11+B2*X21))+
LOG(1/1+PP11)+
(X22*Y2-1/1+NP12)-(B2-a2)/b2**2*(Y2*(B0+B1*X12+B2*X22))+LOG(1/1+PP12)+
(X23*Y3-1/1+NP13)-(B2-a2)/b2**2*(Y3*(B0+B1*X13+B2*X23))+LOG(1/1+PP13)+
(X24*Y4-1/1+NP14)-(B2-a2)/b2**2*(Y4*(B0+B1*X14+B2*X24))+LOG(1/1+PP14)+
(X25*Y5-1/1+NP15)-(B2-a2)/b2**2*(Y5*(B0+B1*X15+B2*X25))+LOG(1/1+PP15)+
(X26*Y6-1/1+NP16)-(B2-a2)/b2**2*(Y6*(B0+B1*X16+B2*X26))+LOG(1/1+PP16)+
(X27*Y7-1/1+NP17)-(B2-a2)/b2**2*(Y7*(B0+B1*X17+B2*X27))+LOG(1/1+PP17)+
(X28*Y8-1/1+NP18)-(B2-a2)/b2**2*(Y8*(B0+B1*X18+B2*X28))+LOG(1/1+PP18)+
(X29*Y9-1/1+NP19)-(B2-a2)/b2**2*(Y9*(B0+B1*X19+B2*X29))+LOG(1/1+PP19)+
(X210*Y10-1/1+NP110)-(B2-a2)/b2**2*(Y10*(B0+B1*X110+B2*X210))+LOG(1/1+
PP110)+(X211*Y11-1/1+NP111)-(B2-a2)/b2**2*(Y11*(B0+B1*X111+B2*X211))+LOG
(1/1+PP111)+(X212*Y12-1/1+NP112)-(B2-a2)/b2**2*(Y12*(B0+B1*X112+B2*X212))+
LOG(1/1+PP112)+(X213*Y13-1/1+NP113)-(B2-a2)/b2**2*(Y13*(B0+B1*X113+B2*
X213))+LOG(1/1+PP113)+(X214*Y14-1/1+NP114)-(B2-a2)/b2**2*(Y14*(B0+ B1*X114
+B2*X214))+LOG(1/1+PP114)+(X215*Y15-1/1+NP115)-(B2-a2)/b2**2*(Y15*(B0+B1*
X115+B2*X215))+LOG(1/1+PP115)+(X216*Y16-1/1+NP116)-(B2-a2)/b2**2*(Y16*
(B0+B1*X116+B2*X216))+LOG(1/1+PP116)+(X217*Y17-1/1+NP117)-(B2-a2)/b2**2*

```
(Y17*(B0+B1*X117+B2*X217))+LOG(1/1+PP117)+(X218*Y18-1/1+NP118)-(B2-a2)
/b2**2*(Y18*(B0+B1*X118+B2*X218))+LOG(1/1+PP118)+(X219*Y19-1/1+NP119)-
(B2-a2)/b2**2*(Y19*(B0+B1*X119+B2*X219))+LOG(1/1+PP119)+(X220*Y20-
1/1+NP120)-(B2-a2)/b2**2*(Y20*(B0+B1*X120+B2*X220))+LOG(1/1+PP120)+
(X221*Y21-1/1+NP121)-(B2-a2)/b2**2*(Y21*(B0+B1*X121+B2*X221))+LOG(1/1+
PP121)+(X222*Y22-1/1+NP122)-(B2-a2)/b2**2*(Y22*(B0+B1*X122+B2*X222))+
LOG(1/1+PP122)+(X223*Y23-1/1+NP123)-(B2-a2)/b2**2*(Y23*(B0+B1*X123+
B2*X223))+LOG(1/1+PP123)+(X224*Y24-1/1+NP124)-(B2-a2)/b2**2*(Y24*(B0+
B1*X124+B2*X224))+LOG(1/1+PP124)+(X225*Y25-1/1+NP125)-(B2-a2)/b2**2*
(Y25*(B0+B1*X125+B2*X225))+LOG(1/1+PP125)+(X226*Y26-1/1+NP126)-(B2-
a2)/b2**2*(Y26*(B0+B1*X126+B2*X226))+LOG(1/1+PP126)+(X227*Y27-1/1+NP127)-
(B2-a2)/b2**2*(Y27*(B0+B1*X127+B2*X227))+LOG(1/1+PP127)+(X228*Y28-1/1+
NP128)-(B2-a2)/b2**2*(Y28*(B0+B1*X128+B2* X228))+LOG(1/1+PP128)+(X229*Y29-
1/1+NP129)-(B2-a2)/b2**2*(Y29*(B0+ B1*X129+B2*X229))+LOG(1/1+PP129)+(X230*
Y30-1/1+NP130)-(B2-a2)/b2**2*(Y30*(B0+B1*X130+B2*X230))+LOG(1/1+PP130)+
(X231*Y31-1/1+NP131)-(B2-a2)/b2**2*(Y31*(B0+B1*X131+B2*X231))+LOG(1/1+
PP131)+(X232*Y32-1/1+NP132)-(B2-a2)/b2**2*(Y32*(B0+B1*X132+B2*X232))+LOG
(1/1+PP132));
   ENDO C0 C1 C2;
   EXO NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
   Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
   X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
   X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131 X132 X21
   X22 X23 X24 X25 X26 X27 X28 X29 X210 X211 X212 X213 X214 X215 X216 X217
   X218 X219 X220 X221 X222 X223 X224 X225 X226 X227 X228 X229 X230 X231
   X232;
   PARMS B0 -1 B1 1 B2 1; RUN;
```

Appendix IV

```
* TO FIND THE POSTERIOR MEAN USING UNIFORM PRIOR OF LOGISTIC
REGRESSION MODEL WITH TWO EXPLANATORY VARIABLES;

Data DD;
aa=0.001;
INPUT NN Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18
Y19 Y20 Y21 Y22 Y23 Y24 Y25 Y26 Y27 Y28 Y29 Y30 Y31 Y32 X11 X12 X13 X14
X15 X16 X17 X18 X19 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119
X120 X121 X122 X123 X124 X125 X126 X127 X128 X129 X130 X131 X132 X21
X22 X23 X24 X25 X26 X27 X28 X29 X210 X211 X212 X213 X214 X215 X216 X217
X218 X219 X220 X221 X222 X223 X224 X225 X226 X227 X228 X229 X230 X231
X232;
CARDS;
32 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
2.52 2.56 2.19 2.18 3.41 2.46 3.22 2.21 3.15 2.60 2.29 2.35 5.06 3.34 2.38 3.15 3.53
2.68 2.60 2.23 2.88 2.65 2.09 2.28 2.67 2.29 2.15 2.54 3.93 3.34 2.99 3.32 38 31 33 31
37 36 38 37 39 41 36 39 37 32 37 36 46 34 38 37 30 46 44 36 39 31 31 28 32 30 36 35
;
run;
Data DDD; set DD;
Do B0= -40 to 40-aa by aa;
Do B1= -7  to  7-aa by aa;
Do B2= -1  to  1-aa by aa;
NP11=EXP(-B0-B1*X11-B2*X21);NP12=EXP(-B0-B1*X12-B2*X22);NP13=EXP(-
B0-B1*X13-B2*X23);NP14=EXP(-B0-B1*X14-B2*X24);NP15=EXP(-B0-B1*X15-
B2*X25);NP16=EXP(-B0-B1*X16-B2*X26);NP17=EXP(-B0-B1*X17-B2*X27);
NP18=EXP(-B0-B1*X18-B2*X28);NP19=EXP(-B0-B1*X19-B2*X29);NP110=EXP(-
B0-B1*X110-B2*X210);NP111=EXP(-B0-B1*X111-B2*X211);NP112=EXP(-B0-
B1*X112-B2*X212);NP113=EXP(-B0-B1*X113-B2*X213);NP114=EXP(-B0-
B1*X114-B2*X214);NP115=EXP(-B0-B1*X115-B2*X215);NP116=EXP(-B0-
B1*X116-B2*X216);NP117=EXP(-B0-B1*X117-B2*X217);NP118=EXP(-B0-
B1*X118-B2*X218);NP119=EXP(-B0-B1*X119-B2*X219);NP120=EXP(-B0-
B1*X120-B2*X220);NP121=EXP(-B0-B1*X121-B2*X221);NP122=EXP(-B0-
B1*X122-B2*X222);NP123=EXP(-B0-B1*X123-B2*X223);NP124=EXP(-B0-
B1*X124-B2*X224);NP125=EXP(-B0-B1*X125-B2*X225);NP126=EXP(-B0-
B1*X126-B2*X226);NP127=EXP(-B0-B1*X127-B2*X227);NP128=EXP(-B0-
B1*X128-B2*X228);NP129=EXP(-B0-B1*X129-B2*X229);NP130=EXP(-B0-
B1*X130-B2*X230);NP131=EXP(-B0-B1*X131-B2*X231);NP132=EXP(-B0-
B1*X132-B2*X232);PP11=EXP(B0+B1*X11+B2*X21);PP12=EXP(B0+B1*X12
+B2*X22);PP13=EXP(B0+B1*X13+B2*X23);PP14=EXP(B0+B1*X14+B2*X24);
PP15=EXP(B0+B1*X15+B2*X25);PP16=EXP(B0+B1*X16+B2*X26);PP17=EXP(B0
+B1*X17+B2*X27);PP18=EXP(B0+B1*X18+B2*X28);PP19=EXP(B0+B1*X19+B2*
X29);PP110=EXP(B0+B1*X110+B2*X210);PP111=EXP(B0+B1*X111+B2*X211);PP
112=EXP(B0+B1*X112+B2*X212);PP113=EXP(B0+B1*X113+B2*X213);PP114=EX
```

```
P(B0+B1*X114+B2*X214);PP115=EXP(B0+B1*X115+B2*X215);PP116=EXP(B0+B
1*X116+B2*X216);PP117=EXP(B0+B1*X117+B2*X217);PP118=EXP(B0+B1*X118
+B2*X218);PP119=EXP(B0+B1*X119+B2*X219);PP120=EXP(B0+B1*X120+B2*X2
20);PP121=EXP(B0+B1*X121+B2*X221);PP122=EXP(B0+B1*X122+B2*X222);PP1
23=EXP(B0+B1*X123+B2*X223);PP124=EXP(B0+B1*X124+B2*X224);
PP125=EXP(B0+B1*X125+B2*X225);PP126=EXP(B0+B1*X126+B2*X226);PP127=
EXP(B0+B1*X127+B2*X227);PP128=EXP(B0+B1*X128+B2*X228);PP129=EXP(B0
+B1*X129+B2*X229);PP130=EXP(B0+B1*X130+B2*X230);PP131=EXP(B0+B1*X1
31+B2*X231);PP132=EXP(B0+B1*X132+B2*X232);
Fun=(1/0.0000034970)*(((1/(1+NP11))**Y1*(1/(1+PP11))**(1-Y1))*((1/(1+NP12))**Y2
*(1/(1+PP12))**(1-Y2))*((1/(1+NP13))**Y3*(1/(1+PP13))**(1-Y3))*((1/(1+NP14))**Y4
*(1/(1+PP14))**(1-Y4))*((1/(1+NP15))**Y5*(1/(1+PP15))**(1-Y5))* ((1/(1+NP16))
**Y6*(1/(1+PP16))**(1-Y6))*((1/(1+NP17))**Y7*(1/(1+PP17))**(1-Y7))*((1/(1+
NP18))**Y8*(1/(1+PP18))**(1-Y8))*((1/(1+NP19))**Y9* (1/(1+PP19))**(1Y9))*
((1/(1+NP110))**Y10*(1/(1+PP110))**(1-Y10))* ((1/(1+NP111))**Y11*(1/(1+PP111
))**(1-Y11))*((1/(1+NP112))**Y12*(1/(1+PP112))**(1-Y12))*((1/(1+NP113))**Y13
*(1/(1+PP113))**(1-Y13))*((1/(1+NP114))**Y14*(1/(1+PP114))**(1-Y14))*((1/(1+
NP115))**Y15*  (1/(1+PP115))**(1-Y15))*((1/(1+NP116))**Y16*(1/(1+PP116))**(1-
Y16))*((1/(1+NP117))**Y17*(1/(1+PP117))**(1-Y17))*((1/(1+NP118))**Y18* (1/(1+
PP118))**(1-Y18))*((1/(1+NP119))**Y19*(1/(1+PP119))**(1-Y19))* ((1/(1+NP120))
**Y20*(1/(1+PP120))**(1-Y20))*((1/(1+NP121))**Y21*(1/(1+PP121))**(1-Y21))*
((1/(1+NP122))**Y22*(1/(1+PP122))**(1-Y22) )*((1/(1+NP123))**Y23*(1/(1+PP123
))**(1-Y23))*((1/(1+NP124))**Y24*(1/(1+PP124))**(1-Y24))*((1/(1+NP125))**Y25
*(1/(1+PP125))**(1-Y25))*((1/(1+NP126))**Y26*(1/(1+PP126))**(1-Y26))*((1/(1+
NP127))**Y27*(1/(1+PP127))**(1-Y27))*((1/(1+NP128))**Y28*(1/(1+PP128))**(1-
Y28))*((1/(1+NP129))**Y29*(1/(1+PP129))**(1-Y29))*((1/(1+NP130))**Y30*
(1/(1+PP130))**(1-Y30))*((1/(1+NP131))**Y31*(1/(1+PP131))**(1-Y31))*
((1/(1+NP132))**Y32*(1/(1+PP132))**(1-Y32)));
fun2=aa**3*fun; integ+fun2;
end;end;end;
proc print data=DDD; var aa integ;
run;
```

References

Aslam, M,. (2003). An application of prior predictive distributions to elicit the prior density. Journal of Statistical Theory and Applications. Vol. 2, No. 1, pp 70-83.

Albert, H. J. & Chib S., (1993). Bayesian Analysis of Binary and Polychotomous Response data. Journal of the American Statistical Association. Vol. 88, No. 422, pp 669-679.

Albert, A., & Anderson, J. A., (1984). On the existence of Maximum Likelihood Estimators in logistic regression models. Biometrica. Vol. 71, pp 1-11.

Amemiya, T., (1981). Quantitative response Models: A survey. Journal of Economic Literature. Vol. 19, pp 1483-1536.

Aldrich, J., (1997). R.A. Fisher and the making of maximum likelihood 1912–1922. Statistical Science. Vol. 12, No. 3, pp 162–176.

Athison, A. C., (1978). Posterior probability for choosing regression model. Biometrika. Vol. 65, No. 1, pp 39-48.

Artis, M., Ayuso, M., Guillen, M., (1999). Modeling different types of automobile insurance fraud behavior in the Spanish market. Insurance: Mathematics & Economics Vol. 34, pp 67-81.

Berger, J. O., (1985). Statistical Decision Theory and Bayesian Analysis, Second Edition New York: Springer-Verlag.

Bernardo, J.M., (1979). Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society *(B)*. Vol. 41, pp 113-147.

Bazan, J., Branco. M., Bolfarine, H., (2006). A skew item response model. Bayesian Analysis. Vol. 1, No. 4, pp 861-892.

Basu, S., Mukhopadhayay, S., (2000). Bayesian analysis of binary regression using symmetric and asymmetric links. Sankhya. Vol. 3, pp 372-387.

Bian, G., (1997). Bayesian Inference in Location-Scale Distributions with Independent Bivariate Priors. Test. Vol. 6, No.1, pp 137-157.

Bian, G., (1989). Bayesian Statistical Analysis with Independent Bivariate Priors for the Normal Location and Scale Parameters. Ph.D. Thesis, University of Minnesota.

Bermudez, L. I., Perez, M. J., Ayuso, M., Gomez, E., & Vazquez, J. F., (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. Journal of Mathematics and Economics. Vol. 42, pp 779-786.

Box, G. E. P., & Tio, G. C., (1973). Bayesian inference in statistical analysis. Reading, Massachusetts: Addition-Wesley.

Bedrick, J. E., Christensen, R., & Johnson, W., (1996). A New Perspective on Priors for Generalized Linear Models. Journal of the American Statistical Association, Vol. 91, No. 436, pp 1450- 1460.

Cengiz A. M. Bek Y. & Yilmaz R., (2001). Bayesian Inference of Binary Logistic Regression Model for Assessing Erythrocyte Sedimentation Rate. Pakistan Journal of Biological Sciences. Vol.4, No.9, pp 1180-1183.

Choi T. Schervish J. M. Schmitt A. K. & Small J. M., (June 2008). A Bayesian Approach to a Logistic Regression Model with Incomplete Information. Biometrics. Vol 64, pp 424-430.

Chaloner, Kathryn. M. and Duncan, George T., (1983). Assessment of Beta prior Distributions: PM Elicitation. The Statistician. Vol. 32, No ½, proceedings of the 1982. I. O. S. Annual conference on practical Bayesian Statistics, pp 174-180.

Chen, M. H, & Ibrahim, J.G., (2003). Conjugate Priors for Generalized Linear Models. Statistica Sinica, Vol. 13, pp 461-476.

Chen, M., Dey, D., Shao, Q., (1999). A new skewed link model for dichotomous quantal response data. Journal of the American Statistical Association Vol. 94, No. 448, pp 1172-1186.

Collett, D., & Jemain, A. A., (1985). Residuals, outliers and influential observations in Regression Analysis. Sains Malaysiana, Vol. 14, pp 493-511.

Collett, D., (1996). Modeling Binary Data. Chapman & Hall.

Crowder M. & Sweeting T., (1989). Bayesian Inference for a Bivariate Binomial Distribution. Biometrika. Vol. 76, No. 3, pp 599-603.

Congdon. P., (2005). Bayesian Methods for Categorical Data, west Sussex, England: John Willy and Sons ltd.

Cox, D. R., (1971). The Analysis of Binary Data. Methuen, London.

Davies, H.T.O., Crombie, I.K., & Tavakoli. M., (1998). When can odds ratios mislead? British Medical Journal Vol.316, pp 989-991.

Denis Conniffe (1997). Improving a Linear Regression Through Joint Estimation with a Probit Model. The Statistician. Vol. 46, No. 4, pp 487-493.

Dreze, J. H., (1977). Bayesian regression analysis using poly-t densities. Journal of Econometrics. Vol. 6, pp 329-354.

Deeks. J., (1996). Swots corner: what is an odds ratio? Bandolier. Vol.3, No 3, Issue 25, pp 6-7.

El-Sayyad, G. M., (1973) Bayesian and classical analysis of Poisson Regression. Journal of Royal Statistical Society. Series B (Methodological). Vol. 35, No 3, pp 445-451.

Easves, D. M., & Chang, T., (1992). Posterior mode estimation for the generalized linear model. Annals of the Institute of Statistical Mathematics. Vol. 44, pp 417-434.

Gelman, A., Jakulin, A., Pittau, G. M., & Su, S. Y., (2008). A weekly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. Vol. 2, pp 1360-1383.

Geweke, J., (1991). Efficient simulation from the multivariate normal and student-t distribution subject to linear constraints. Computer science and Statistics: proceedings of the Twenty-third symposium on the interface, 571-578.

Ghosh, J. K., & Mukerjee, R., (1992). Non-informative priors. In *Bayesian Statistics*. Vol. 4, pp 195–210. Oxford Univ. Press, New York.

Geweke, J., Keane, M., (2001). Computationally intensive methods for integration in econometrics. In Hechman, J., Leamer, E., Hand book of econometrics. Chap. 5, Elsevier, pp 3463-3568.

Griffiths, W. E., Hill, R. C., Pope, P. J., (1987). Small sample properties of probit model estimators. Journal of American Statistical Association. Vol. 82, pp 929-937.

Groenewald, P. C. N., & Mokgatlhe, L., (2005). Bayesian computation for logistic regression. Computational Statistics & Data Analysis. Vol. 48. pp 857-868.

Hausman, J.A., Wise D.A., (1978). A conditional probit model for quantitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica. Vol. 46, pp 407-426.

Hausman, J., McFadden, D., (1984). Specification tests for the multinomial logit model. Econometrica. Vol. 52, No. 5, pp 1219-1240.

Hoffmann, P. J., (2004). Generalized linear models, an applied approach. First edition. Boston: Pearson Education, Inc.

Hajivassilious, V., (1990). Smooth simulation estimation of panel data. LDV models. Mimeo. Yale University.

Ibrahim, Joseph G. and Laud, Purushottam W., (1991). On Bayesian analysis of generalized linear models using Jeffrey's prior. Journal of American Statistical Association. Vol. 89, No. 416, pp 981-986.

Ibrahim, Joseph G., (1997). On properties of predictive prior and linear models. The American Statistician. Vol 51, No. 4, pp 333-337.

Ibrahim, J., & Loud, P., (1991). On Bayesian analysis of generalized linear models using Jeffreys Prior. Journal of American Statistical Association. Vol. 86, pp 981-986.

Jaynes, T. E., (1968). "Prior Probabilities", IEEE Transactions on Systems Science and Cybernetics, SSC-4, 227-241, Reprinted in Rosenkrantz, D. R., Compiler & Jaynes, T. E., (1983). Papers on Probability, Statistics and Statistical Physics. Dordrecht, Holland: Reidel Publishing Company, pp 116-130.

Jeffreys, H., (1961). Theory of probability. Oxford University Press, Oxford, U.K.

John, W. P., (1976). Fisher, R. A., Edgeworth, F. Y., On the Efficiency of Maximum Likelihood Estimation. The Annals of Statistics. Vol. 4, No. 3, pp 501–514.

Keane, M., (1994). A computationally practical simulation estimator for panel data. Econometrica. Vol. 62, pp 95-116.

Kass, R.E., & Wasserman, L., (1996). The selection of prior distributions by formal rules. Journal of the American Statistical Association. Vol. 91, pp 1343-1370.

King, E. N., and Ryan, T. P. (2002), A Preliminary Investigation of Maximum Likelihood Logistic Regression Vs Exact logistic regression. The American statistician. Vol. 56, pp 163-170.

Koop, G., Poirier, D., (1993). Bayesian analysis of logit models using natural conjugate priors, Journal of Econometrics. Vol. 56, No. 3, pp 323-340.

Liesenfeld, R., & Richard, F. J., (2009). Efficient estimation of Probit models with correlated errors. Journal of Econometrics. Vol. 156, No. 2, pp 367-376.

Laplace, P. S. (1812). *Théorie Analytique des Probabilités*. Paris: Courcier. Reprinted as, *Oeuvres Complètes de Laplace*. Vol. 7, pp 1878–1912. Paris: Gauthier-Villars.

Maddala, G. S., (1983). Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, New York.

McCulloch, R., Polson, N., Rossi, P., (1999). A Bayesian analysis of the multinomial probit model with fully identified parameters. Journal of Econometrics. Vol. 99, No. 1 pp 173-193.

McFadden, D., (1981). Econometric models of probabilistic choice. In: man-ski, C., McFadden, D., (Eds), Structural Analysis of Discrete Data. MIT press, Cambridge, pp 198-272.

Munkin K. M. & Trivedi K. P., (2008). Bayesian Analysis of the Ordered Probit Model with Endogenous Selection. Journal of Econometrics. Vol. 143, pp 334-348.

Poirier D., (1994). Jeffrey's prior for Logit Models. Journal of Econometrics. Vol. 63, pp 327-339.

Rossi, E. P., (1996). Existence of Bayes Estimators for the Binomial Logit Model. In Berry, D. A., Chaloner, K. M., & Geweke, K. J., Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner. New York: John Wiley, pp. 91-100.

Rijmen, F., (2008). Bayesian networks with a logistic regression model for the conditional probabilities. International Journal of Approximate Reasoning. Vol. 48. pp 659-666.

Rubin, D. B., (1981). The Bayesian Bootstrap. Annals of Statistics, Vol. 9, No. 1, pp 130-134.

Silvapulle, J. M., (1981). On the existence of Maximum Likelihood Estimators for the Binomial response models. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 43, No. 3, pp 310-313.

Smous, Evan P., (1984). A Note on Bayesian Least Square Inference for Finite Population Models, Journal of the American Statistical Association. Vol 79, No. 386, pp 390-392.

Smith M. F. A. (1973). A General Bayesian Linear Model. Journal of the Royal Statistical Society. Series B (methodological). Vol. 35, No.1, pp 67-75.

Stukel, T., (1988). Generalized logistic model. Journal of the American Statistical Association Vol. 83, pp 426-431.

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

Tekas D. & Gunay S., (2008). A Bayesian Approach to Parameter Estimation in Binary Logit and Probit Models. Hacettepe Journal of Mathematics and Statistics. Vol. 37, No.2, pp 167-176.

Thurstone, L., (1927). A law of comparative judgment. Journal of Psychological Review. Vol. 34, pp 273-286.

Tibshirani, R., (1989). Noninformative priors for one parameter of many. Biometrika. Vol. 76, N0.3, pp 604-608.