BAYESIAN DATA ANALYSIS USING STATISTICAL SOFTWARES

Dissertation Submitted

In partial fulfillment for the award of degree of Master of Philosophy

in

STATISTICS

By

RAJA SULTAN AHMAD RESHI

Under the Supervision of

Dr. SHEIKH PARVAIZ AHMAD

Assistant Professor

DEPARTMENT OF STATISTICS

UNIVERSITY OF KASHMIR

Hazratbal, Srinagar, J&K, India

NAAC Accredited Grade "A"

2012

DEPARTMENT OF STATISTICS

University of Kashmir,
Hazratbal, Srinagar 190006

Dr. Sheikh Parvaiz Ahmad
(Ph.D., UoK)
Assistant Professor
Web Add:kashmiruniversity.net

E-mail: sprvz@yahoo.com
Mob.9906678065
Telefax:0194-2421357

Dated: ______________

CERTIFICATE

This is to certify that the scholar **Mr. Raja Sultan Ahmad Reshi**, has carried out the present dissertation entitled **"Bayesian Data Analysis Using Statistical Softwares"** under my supervision and the work is suitable for submission for the award of degree of Master of Philosophy in Statistics. It is further certified that the work has not been submitted in part or full for the award of M.Phil or any other degree.

Dr. Sheikh Parvaiz Ahmad
Supervisor

ACKNOWLEDGEMENT

It brings me colossal pleasure while placing on record my sincere gratitude and reverence to my esteemed supervisor, Dr. Sheikh Parvaiz Ahmad whose diligent supervision, intellectual guidance, deep concern and wisdom natured the growth and defined the direction of my study.

Words fail to express my indebtedness and gratitude to Professor Aquil Ahmad (Head of Deptt. Statistics) for his encouragement, undaunted support and other invaluable help, without which, this task could not have been accomplished.

I feel deeply overwhelmed in expressing my sincere regards and wish to extend my heartfelt thanks to Dr.Anwar Hassan, Dr. M. A. K. Baig and Dr. Tariq Rashid Jan for their moral support and golden words of encouragements during the entire course of my study.

Completion of this work was only possible due to the cooperation and help of many people. I shall be failing in my duty if I do not record my indebtedness to my fellow research scholar particularly Adil Khan, Hummara sultan, Shabir Ahmad, kaisar Ahmad, kounsar, Nusrat Mushtaq, Gulzar Ahmad Javid Ahmad, Bilal Ahmad and my dearest friend Adil Rashid, who constantly helped and gave me the valuable suggestions I need at many times.

I would also like to express my gratitude towards the library staff for providing me access to the study material available in our Department library. Credit also goes to the technical staff of the P.G. Department of the Statistics for their assistance and help.

I consider it to be a big mistake if I forget to express my gratitude to those who have groomed me throughout my life. I will be failing in my duty if I ignore the constant help rendered by my parents, *my sister and my near and dear ones while pursuing the work. My family deserves all the credits and accolades earned by the achievement of this distinction. My heartfelt devotion is due to my friends during my PG programme especially to Fayaz Ahmad, Zahoor Ahmad, Arshid, Tanveer, Khalid and Tanveer (CR) who always were there as my true energy and strength.*

Raja Sultan Ahmad Reshi

Dedicated to…

My

Beloved

Grand Parents

&

Parents

CONTENTS

Preface:

Bayesian data analysis is practical method for making inferences from data using probability models for quantities we observe and for quantities about we wish to learn. Bayesian approach is an excellent alternative to use large sample procedure and is likely to be more reasonable for moderate and especially small sample sizes where non Bayesian procedures break down. Bayesian data analysis combines Bayesian probability theory with statistical data analysis to make predictions about future events based on our current information. Source of information from data is summarized in the form of likelihood while that of non data is termed as prior information. Posterior density is the final outcome after combining these two sources of information. In this thesis we have tried to construct posterior distributions, with its practical applications. The thesis is divided into five chapters:

Chapter I includes introduction to Bayesian data analysis, Bayes theorem, sequential nature of Bayes theorem, likelihood, marginal and posterior distribution, predictive distribution, highest posterior density and some important probability models like Binomial, Poisson and Normal distributions.

Chapter II is devoted to the introduction of prior and different types of prior. Some important loss functions such as squared- error loss function, weighted SELF, quadratic SELF, linear loss, absolute loss, zero-one loss, risk function are discussed. Estimation techniques and large sample approximations like Laplace, Lindely and normal approximations are also discussed.

Chapter III deals with the Bayesian analysis of parameters of binomial distribution under different priors. Normal, Lindely's approximation to the posterior density of binomial distribution are also discussed. Some programmes in S-PLUS and R softwares have been developed for numerical and graphical representation of posterior densities and credible interval under different priors.

Chapter IV deals with the Bayesian estimation of Poisson distribution under different priors, comparisons of different priors are done with respect to posterior variance, Bayesian point estimates, using coefficient of skewness. We have also discussed posterior distribution under different double priors. Computer programmes in R-software are developed to illustrate numerical data.

Chapter V is completely devoted to the Bayesian analysis of normal distribution. This chapter contains Bayesian estimator and Credible intervals for parameters of normal distribution, the posterior distribution and the posterior predictive distribution for the unknown parameter σ^2 is discussed using different type of prior distribution. Methods proposed in this chapter are illustrated numerically in R-software.

CHAPTER – 1

INTRODUCTION

TO

BAYESIAN DATA ANALYSIS

1.1 Introduction:

Experiments are performed commonly in three steps; first, the experiment must be designed; second, the data must be gathered; and third, the data must be analyzed. These three steps are highly idealized, and no clear boundary exists between them. The problem of analyzing the data is one that should be faced early in the design phase. Gathering the data in such a way as to learn the most about a phenomenon is what an experiment is all about.

In many experiments it is essential that one should does the best possible job in analyzing the data. This could be true because no more data can be obtained, or one is trying to discover a very small effect. Furthermore, thanks to modern computers, sophisticated data analysis is far less costly than data acquisition, so there is no excuse for doing the best job of analysis that one can.

By Bayesian data analysis, we mean practical methods for making inferences from data using probability models for quantities we observe and for quantities about which we wish to learn. The essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty in inferences on statistical data analysis.

The process of Bayesian data analysis can be idealized by dividing it into three steps:

a) Setting up a full probability model – a joint probability distribution for all observable and unobservable quantities in a problem. The model should be consistent with knowledge about the underlying scientific problem and the data collection process.
b) Conditioning on observed data: Calculating and interpreting the appropriate posterior distribution-the conditional probability distribution of unobserved quantities of ultimate interest, given the observed data.
c) Evaluating the fit of model and implications of the resulting posterior distributions.

Great advances in all these areas have been made in the last twenty-five years. A primary motivation for believing Bayesian thinking is that it facilitates a common sense interpretation of statistical conclusions.

Bayesian statistics requires the mathematics of probability theory and the interpretation of probability which most closely corresponds to the standard use of this word in everyday language: it is no accident that some of the more important seminal books on Bayesian statistics such as the works of the Laplace (1812), Jefferys (1939) and De Finetti (1970) are actually entitled “probability theory”. Bayesian approach to statistics is very different from the classical methodology, it formally seeks use of prior information and Bayes theorem provides the basis for making use of this information. When significant prior information is available, the Bayesian approach shows how to utilize it sensibly. This is not possible with the most non-Bayesian approaches. The business of statistics is to provide information or conclusions about uncertain quantities. The language of uncertainty is probability. Bayesian approach consistently uses this language to directly address uncertainty. The classical or frequentists interpret probability as the limit of the success ratio as the number of trails ‘n’ conceptually tends to infinity. Under this interpretation the parameter θ in a statistical model is treated as an unknown constant and the sample of observations $Y=(y_1, y_2, \ldots, y_n)$ is regarded as the random sample from some underlying distribution. The classical school believes in Fishers Likelihood Principle which claims that all the information about the unknown parameter(s) is contained in the sample as summarized by the likelihood function. This Principle leads to Fishers maximum likelihood estimator.

On the other hand for Bayesian approach probability is a person’s degree of belief in a certain proposition ‘A’ based on the prior (or current) knowledge about A and this degree of belief is successively revised or updated as new information is available about the proportion. In Bayesian framework the parameter is justifiably regarded as a random variable and the data once obtained is given or fixed. For example, in the exponential model the mean life θ may be regarded as varying from batch to batch overtime and this variation is represented by a probability distribution over parameter space Ω. Thus the basic difference in the two approaches may be explained in the single sentence that to a frequentist, the parameter is constant and he is suspicious about the data, where as to a Bayesian data is given (or fixed) and he is suspicious about the parameter. Bayesian approach is an excellent alternative to use large sample procedure and is likely to be more reasonable for moderate and especially small sample sizes where non Bayesian procedures break down (e.g., Berger, 1985).

In the Bayesian framework we assign degrees of beliefs for different events the approach is also called the subjective probability opposed to objective probability approach that the frequentists use. The difference between these models is easily illustrated by an example from real life. Consider that two football teams A and B are playing against each other. What kind of probability could we assign to event ‘A’ win? The Bayesian would assign his own subjective probability (belief) to the event where as the frequentist would make statistical analysis about the games these teams have played against each other.

Bayesian data analysis combines Bayesian probability theory with statistical data analysis to make predictions about future events based on our current information. Bayesian data analysis can also be defined as a practical method for making inference from data using probability models for quantities we observe and for quantities about which we wish to learn by sitting up a full probability model, conditioning on observed data and evaluating the fit of the model e.g. we can make a prediction whether team B is going to win or not using Bayesian approach. The probability is encoded in the model which contains all relevant observable and unobservable (latent) quantities and this model is then fitted to the available data. After this part we make predictions about future events based on our model.

1.2 Bayes Theorem:

Bayes theorem is an essential element of the Bayesian approach to statistical inference. The central feature of Bayesian inference is the direct quantification of uncertainty in terms of probabilistic statements. Often, we begin our analysis with initial or prior probability estimates for specific events of interest then, from sources such as a sample, a special report, a product test and so on we obtain some additional information about the events. Given this new information we update the prior probability values by calculating revised probabilities, referred to as posterior probabilities. The steps in this probability revision process are shown in the following diagram.

The origin of Bayes theorem has a fascinating history. It is named after the Rev. Thomas Bayes, a priest who never published a mathematical paper in his lifetime. The paper in which the theorem appears was posthumously read before the royal society by his friend Richard Price in 1764. Stigler suggests it was first discovered by Nicolas Saunderson, a blind mathematician/optician who, at age 29, became lucasian Professor of mathematics at Cambridge (the position earlier held by Issac Newton). More details are discussed in Stigler, 1983.

a) Bayes Theorem for Events:

The probability of an event 'A' depends upon the available information about the event 'A'. For example, if we have a die having two faces with the number 6 and if the event 'A' is that any number other than 6 appears on the die then p(A)=2/3 and it isn't 5/6 (when the die is considered a fair one and having distinct numbers on its faces). In order to represent the prior information that the die had two faces with the number 6 and denote the event 'B' then we should have used the notation $P(A|B)$ instead of p(A).

Bayes theorem is the basic rule for incorporating the prior information that the event 'B' has occurred and influences evaluation of the probability for the event 'A'. The simplest form of the Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}; P(B) > 0$$

Follows easily from the definition of conditional probability

$$P(A)P(A|B) = P(A \cup B) = P(B)P(A|B)$$

It provides a mechanism of the process of learning by experience. The connection between $P(A|B)$ and $P(B|A)$ together with the initial probability P(A) is the basis for the process of acquiring knowledge. In general given two events A and B, the inductive reasoning consists in applying Bayes theorem which answers how the information about the occurrence of event B influences P(A). The posterior probability $P(A|B)$ is proportional to the initial (prior) probability P(A)and the so called likelihood $\mathrm{P(B|A)}$. This is the process by which we learn from experience in the sense that experience gives us information that can modify our initial belief according to the factor $P(B|A)/P(B)$.

b) Generalized Bayes Theorem For Events:

It states that, if $A_1, A_2, ..., A_k$ are k mutually exclusive and exhaustive events and B is another independent event such that $P(B | A_i)$ is the conditional probability of B given that A_i has already occurred, then

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{i=1}^{k} P(B | A_i)P(A_i)} : i = 1,2,...,k$$

Proof: Let $A_1, A_2, .., A_k$ constitutes a partition of the sample space S

i.e., $S = A_1 \cap A_2 \cap ... A_k \ and \ A_i \cap A_j = \phi ; i \neq j$

The events $A_1, A_2, ..., A_k$ are mutually exclusive and exhaustive events (since the union of the disjoint sets equal to the sample space S).

Furthermore, suppose the prior probability of the event A_i is positive

i.e.: $P(A_i) > 0; \forall i = 1,2,...., k$

Suppose an event B can occur only if one of the mutually exclusive and exhaustive events $(A_1, A_2, ..., A_k)$ occurs

We have by definition of conditional probability

$$P(B | A_i) = \frac{P(B \cap A_i)}{P(A_i)}$$

$$P(B \cap A_i) = P(A_i)P(B | A_i)$$

and $P(A_i \cap B) = P(B)\, P(A_i | B)$

but $P(B \cap A_i) = P(A_i \cap B)$

or $$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{P(B)} \qquad (1.2.1)$$

Since 'B' can occur only if A_1 or A_2 or …….A_k occurs it follows that 'B' is the union of 'k' mutually exclusive events

i.e., $B = (B \cap A_1) \cap (B \cap A_2) \cap .... \cap (B \cap A_k)$

Since $A_1, A_2, ..., A_k$ are mutually exclusive and exhaustible events

$\therefore P(B) = P(B \cap A_1) + P(B \cap A_2) + ... + P(B \cap A_k)$

or $P(B)=\sum_{i=1}^{k} P(B \cap A_i)$ (1.2.2)

Using equation (1.2.1) and (1.2.2) we obtain

$$P(A_i \mid B)=\frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{k} P(B \mid A_i)P(A_i)}$$

The conditional probability $P(A_i \mid B)$ is often called the posterior probability because it represents a probability computed after the sample information is taken into account i.e. a probability which has undergone revision via Bayes rule. The probability before revision by Bayes rule is called prior probability.

c) Bayes Theorem for Random Variables:

Suppose that $Y^T = (y_1, y_2, \ldots\ldots, y_n)$ is a vector of n observations whose probability distribution $P(Y \mid \theta)$ depends upon the value of k Parameter $\theta^T = (\theta_1, \theta_2, \ldots\ldots, \theta_k)$. Suppose that θ itself has a probability distribution $P(\theta)$. Then

$$P(Y \mid \theta)P(\theta) = P(Y, \theta) = P(\theta \mid Y)P(Y) \tag{1.2.3}$$

Given the observed data Y, the conditional distribution of θ is

$$P(\theta \mid Y) = \frac{P(Y \mid \theta)P(\theta)}{P(Y)} \tag{1.2.4}$$

Also we can write

$$P(Y) = E[P(Y \mid \theta)] = k^{-1} = \int P(Y \mid \theta)P(\theta)\, d\theta; \qquad \theta \textit{ continuous}$$

$$\sum P(Y \mid \theta)P(\theta); \qquad \theta \textit{ discrete}$$

where the sum or the integral is taken over the admissible range of θ, and where $E[f(\theta)]$ is the mathematical expectation of $f(\theta)$ with respect to the distribution $P(\theta)$. Thus we may write (1.2.4) alternatively as

$$P(\theta \mid Y) = k\ P(Y \mid \theta)P(\theta) \tag{1.2.5}$$

The statement of (1.2.4) or its equivalent (1.2.5) is usually referred to as Bayes theorem. In this expression, $P(\theta)$ which tells us what is known about θ without knowledge of data, is called prior distribution of θ, or the distribution of θ a priori. The density $P(Y \mid \theta)$ is likelihood function of θ which represents the contribution of Y(data) to knowledge about θ (e.g., Berger,1985 and Zellner, 1971). Finally, $P(\theta \mid Y)$, which tells us what is known about given knowledge of the data, is called the posterior distribution of θ given Y, or the distribution of θ a posteriori. The quantity k is merely a "normalizing" constant necessary to ensure that the posterior distribution $P(\theta | Y)$ integrates or sums to one.

1.3 Sequential Nature of Bayes Theorem:

Now given the data Y, $P(Y|\theta)$ in (1.2.5) may be regarded as a function not of Y but of θ. When so regarded, following Fisher (1922), it is called the likelihood function of θ for given Y and can be written as $L(\theta | Y)$.We can thus write Bayes formula as

$$P(\theta | Y) = L(\theta | Y)P(\theta) \qquad (1.3.1)$$

The theorem in (1.3.1) is appealing because it provides a mathematical formulation of how previous knowledge may be combined with new knowledge. Indeed the theorem allows us to continually update information about a set of parameters θ as more observations are taken. Thus, suppose we have an initial sample of observations Y_1, then Bayes initial formula gives.

$$P(\theta | Y_1) \propto P(\theta)L(\theta | Y_1) \qquad (1.3.2)$$

Now suppose we have a second sample of observation Y_2, distribution independently of first sample, then

$$\begin{aligned} P(\theta | Y_1, Y_2) &\propto P(\theta) L(\theta | Y_1) L(\theta | Y_2) \\ &\propto P(\theta | Y_1) L(\theta | Y_2) \qquad (1.3.3) \end{aligned}$$

The expression (1.3.3) is precisely of the same form as (1.3.2) except that $P(\theta | Y_1)$, the posterior distribution for θ given Y_1, plays the role of the prior distribution for the second sample. Obviously this process can be repeated any number of times. In particular, if we have n independent observations the posterior distribution can, if desired, be recalculated after each new observation, so that at the m^{th} stage the likelihood associated with the m^{th} observation is combined with the posterior distribution of θ after m-1 observations to give the new posterior distribution.

$$P(\theta | Y_1, Y_2, ..., Y_m) \propto P(\theta | Y_1, Y_2, ..., Y_{m-1}) L(\theta | Y_m) : m = 2, ..., n \qquad (1.3.4)$$

where $P(\theta | Y_1) \propto P(\theta)L(\theta | Y_1)$

Thus, Bayes theorem describes in a fundamental way, the process of learning from experience and shows how knowledge about the state of nature represented by θ is continually modified as new data becomes available (e.g. Box and Tiao, 1973).

1.4 From Likelihood to Bayesian Analysis:

The method of maximum likelihood and Bayesian analysis are closely related. Suppose $L(\theta \mid Y)$ is the assumed likelihood function. Under ML estimation, we would compute the mode (the maximal value of L, as a function of θ given the data Y) of the likelihood function and use the local curvature to construct the confidence intervals. Hypothesis testing follows using likelihood ratio (LR) statistics. The strength of ML estimation rely on its large –sample properties, namely that when the sample size is sufficiently large, we can assume both normality of the test statistic about its mean and that LR tests follows χ^2 distributions. These nice features don't necessarily hold for small samples (e.g, Gianola and Fernando, 1986).

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

$$P(\theta \mid Y) = \frac{1}{P(Y)} P(Y \mid \theta) P(\theta) \qquad (1.4.1)$$

$$= \text{(normalizing constant) } P(Y \mid \theta) P(\theta) \qquad (1.4.2)$$

$$= \text{constant .likelihood .prior} \qquad (1.4.3)$$

As $P(Y \mid \theta) = L(\theta \mid Y)$ is just the likelihood function. 1/P(Y) is constant (with respect to θ), because our concern is the distribution over θ. Because of this, the posterior is often written as

$$P(\theta \mid Y) \propto L(\theta \mid Y) P(\theta) \qquad (1.4.4)$$

where the symbol $\propto$ means "proportional to" (equal up to a constant). Note that the constant P(Y) normalizes $P(Y \mid \theta)P(\theta)$ to one, and hence can be obtained by integration

$$P(Y) = \int_{\theta} P(Y \mid \theta) P(\theta) d\theta \qquad (1.4.5)$$

The dependence of the posterior on the prior (which can easily be assessed by trying different prior) provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected under different priors, the data are likely highly informative. To see this taking logs on equation (1.4.4) (and ignoring the normalizing constant) gives

Log (posterior) =log (likelihood) +log (prior) (1.4.6)

The Standard Likelihood

When the integral $L(\theta \mid Y)d\theta$ taken over the admissible range of θ is finite, then occasionally it will be convenient to refer to the quantity

$$\frac{l(\theta \mid Y)}{\int l(\theta \mid Y)d\theta} \tag{1.4.7}$$

We shall call this the standardized likelihood that is the likelihood scaled so that the area, volume or hyper volume under the curve, surface or hyper surface is one.

1.5 Marginal posterior distribution:

Often only a subset of the unknown parameter is really of concern to us, the rest being nuisance parameter that are of no concern to us. A very strong feature of Bayesian analysis is that we can remove the effect of nuisance parameters by simply integrating them out of the posterior distribution to generate a marginal posterior distribution for the parameters of interest. For example, suppose the mean and variance of the data coming from a normal distribution are unknown, but our real interest is in the variance. Estimating the mean introduces additional uncertainty into our variance estimate. This is not fully captured in standard classical approaches but under Bayesian analysis, the posterior marginal distribution for σ^2 is simply

$$P(\sigma^2 \mid Y) = \int p(\mu, \sigma^2 \mid y)d\mu$$

The marginal posterior may involve several parameters (generating joint marginal posteriors). Write the vectors of unknown parameters as $\theta = (\theta_1, \theta_n)$ where θ_n is the vector of nuisance parameters. Integrating over θ_n gives the desire marginal as

$$P(\theta_1 \mid Y) = \int_{\theta_n} p(\theta_1, \theta_n \mid y)d\theta_n$$

The requirement of orthogonality between nuisance parameters and the parameters of interest is not required in this framework (e.g., Cox and Reid, 1987). Moreover, marginal posterior densities are better substitutes of conditional profile likelihoods, of Cox and Reid (1987).

1.6 Summarizing the posterior distribution:

If our mindset is to use some sort of point estimator (as is usually done in classical statistics); there are a number of candidates, we could follow maximum likelihood and use the mode of the distribution (its maximal value), with $\hat{\theta} = \max_{\theta}[p(\theta \mid y)]$. We could take the expected value of θ given the posterior

$$\hat{\theta} = E[\theta \mid Y] = \int \theta P(\theta \mid y)\, d\theta$$

Another candidate is the median of the posterior distribution, where the estimator satisfies $\Pr(\theta > \hat{\theta} \mid Y) = \Pr(\theta < \hat{\theta} \mid Y) = 0.5$

and hence

$$\int_{\hat{\theta}}^{\infty} P(\theta \mid Y) d\theta = \int_{-\infty}^{\hat{\theta}} P(\theta \mid Y) d\theta = 1/2$$

However, using any of the above estimators or even all three simultaneously loses the full power of the Bayesian analysis, as the full estimator is the entire posterior density itself. If we cannot obtain the full form of the posterior distribution, it may still be possible to obtain one of the three above estimators.

It is to be noted that under the squared loss function, the Bayesian estimator of θ is defined as the posterior expectation of θ**,** given the data $Y = (y_1, y_{2,} ..., y_n)'$

$$E(\theta \mid Y) = \int_{\Omega} \theta P(\theta \mid Y) d\theta$$

and under weighted squared loss function, the Bayesian estimator θ given data by $Y = (y_1, y_{2,} ..., y_n)'$ is given by

$$\mathrm{a} = \frac{\mathrm{E}(\theta\,\omega(\theta) \mid \mathrm{Y})}{\mathrm{E}(\omega(\theta) \mid \mathrm{Y})},$$

provided the expectations exists. Where $\omega(\theta)$ is the weight associated with loss function.

1.7 Predictive Distribution:

It is the pdf (or pmf) of the as yet unobserved observation x given sample information Y. let us write $f(x, \theta \mid Y) = f(x \mid \theta, Y) P(\theta \mid Y)$ as the joint pdf of X and the parameter Θ, given the sample information Y. Here $f(x \mid \theta, Y)$ is the conditional pdf for x given Θ and Y, where $P(\theta \mid Y)$ is the conditional pdf for Θ given Y the predictor pdf $f(x \mid Y)$ is obtained as:

$$f(x \mid Y) = \int f(x, \theta \mid Y)\, d\theta = \int f(x \mid \theta, Y)\, p(\theta \mid Y)\, d\theta$$

In case, the unobserved observation of x is independent of sample information Y, that is x and y have independent conditional pdf's then

$$f(x|Y) = \int f(x|\theta)p(\theta|Y)\, d\theta$$

1.8 Bayes Rule:

We have some prior information suggesting, that some values of θ are more probable than other. Then the average risk associated with d, with respect to prior distribution $P(\theta)$ is

$$r(g,d) = \int_{\Theta} R(\theta,d)P(\theta)d\theta = \int_{\Theta}\left[\int_{\Omega} L(\theta,d)dF(\theta)\right]P(\theta)d\theta$$

and Bayes rule suggests choosing that d for which r(g,d) is minimum i.e Bayes rule states that choose $\mathrm{d}^* \in \mathrm{D}$ if

$$r(g,d^*) \le r(g,d); \forall d \in D$$

A decision function d which minimizes r(g,d) is called Bayes solution of the decision problem w.r.t the prior density $\mathrm{P}(\theta)$.

The resulting minimum of r(g,d) is called a Bayes risk relative to P(.). In order to apply the Bayes rule, we have to assume that the elements of Θ are the values of random variable $\hat{\theta}$, whose density $\mathrm{P}(\theta)$ is known. In the problem of the estimation of a parameter Θ, with the loss function proportional to the squared error, Bayes rule w.r.t. a given prior distribution is to estimate θ by posterior mean.

To fix the idea, let $f(y|\theta)$ be the pdf of Y and $\mathrm{P}(\theta)$ be the prior density of θ the joint density of Y and θ is

$$h(Y,\theta) = P(\theta)\, f(y|\theta)$$

Hence the posterior distribution of θ given Y=y, has the density

$$P(\theta|Y) = \frac{f(y|\theta)P(\theta)}{\int_{\Theta} f(y|\theta)P(\theta)d\theta}$$

Thus when the loss function is proportional to the squared error, the posterior expected loss, given Y=y is

$$\int_{\Theta} L(\theta,d)P(\theta|y)d\theta$$

To find the action d that minimize this expression, we may set the derivative w.r.t. d in $\int_{\Theta} L(\theta, d) P(\theta \mid y) d\theta$ equal to zero, i.e.

$$\int_{\Theta} \frac{\partial}{\partial d} L(\theta, d) P(\theta \mid y) d\theta = 0$$

The solution will be Bayes estimator. It may be noted that if loss function is squared error, a Bayesian decision rule w.r.t. a given prior distribution of θ is to estimate the mean of the posterior distribution of θ, given the observation.

1.9 Highest posterior density (HPD):

Once the posterior distribution $P(\theta|Y)$ is obtained we may ask, "How likely is it that θ lies within a specified interval $[c_1, c_2]$?" This is not the same as the classical confidence interval interpretation for θ. Since θ is a constant and it is meaningless to make a probability statement about a constant. Posterior intervals based on non informative priors were called credible intervals by Edwards, Lindeman and Savage (1963) and Bayesian confidence intervals by Lindley (1965). It is an interval which contains a certain fraction of the degree of belief. The interval $[c_1, c_2]$ is said to be a $(1-\alpha)$ credible interval for θ if

$$\int_{c_1}^{c_2} P(\theta \mid Y) d\theta = 1 - \alpha \qquad (1.9.1)$$

For the shortest credible interval we have to minimize $I = c_2 - c_1$ subject to the condition (1.9.1) which requires

$$P(c_1 \mid Y) = P(c_2 \mid Y) \qquad (1.9.2)$$

An interval $[c_1, c_2]$ which simultaneously satisfies (1.9.1) and (1.9.2) is called the shortest $(1-\alpha)$ credible interval. An equal tail $(1-\alpha)$ credible interval $[c_1, c_2]$ for θ is given by

$$\int_{-\infty}^{c_1} P(\theta \mid Y) d\theta = \int_{c_2}^{\infty} P(\theta \mid Y) d\theta = \frac{\alpha}{2}$$

An interval which simultaneously satisfies the following conditions:

i) For a given probability content, P the interval should be as short as possible.

ii) The posterior density at every point inside the interval be greater than that for every point outside it so that the interval includes more probable values of the parameter and excludes less probable values is called the Highest posterior Density or HPD – interval (Box and Tio,1973).

For a unimodal but not necessarily symmetrical posterior density, the shortest credible and the HPD-intervals are one and same. (Evans, 1976). However the situation becomes more complicated for the multimodal posterior distribution. The highest mode is determined and the HPD-credible interval is constructed around it. The obtained HPD interval may not be unique. It may happen that the HPD region is a union of two disjoint intervals. Such situations occur when we consider posterior distributions obtained from mixtures of prior densities. An approximation to HPD credible interval may be obtained through the use of the normal approximation of the posterior distribution.

It is critical to note that there is profound difference between a confidence interval (CI) from classical (frequentists) statistics and a Bayesian interval. The interpretation of a classical confidence interval is that if we repeat the experiment a large number of times and construct CIs in the same fashion, that $(1-\alpha)$ of times the confidence interval will enclose the (unknown) parameter. With a Bayesian confidence interval, there is a $(1-\alpha)$ probability that the interval contains the true value of the unknown parameter. Often the CI and the Bayesian intervals have essentially the same value, but again the interpretational difference remains. The key point is that the Bayesian prior allows us to make direct probability statements about θ, where under classical statistics we can only make statements about the behavior of the statistic if we repeat an experiment a large number of times.

1.10 Some Important Probability Distributions:

i)Binomial distribution: This distribution is also known as the Bernoulli distribution after the Swiss mathematician James Bernoulli (1654-1705) who discovered it in 1700 and was first published in 1713, eight years after his death. This distribution can be used under the following conditions:

i) The random experiment is performed repeatedly a finite and fixed of times. In other words n, the number of trials is finite and fixed.

ii) The outcome of the random experiment (trial) results in the dichotomous classification of events. In other words, the outcome of the trial may be classified into two mutually disjoint categories called success (the occurrence of the event) and failure (the non-occurrence of event)

iii) All the trials are independent, i.e. the result of any trial, is not affected in any way by the preceding trials and does not affect the result of succeeding trials.

iv) The probability of success (happening of event) in any trial is θ and is constant for each trial. $1-\theta$ Is then termed as the probability of failure (non-occurrence of the event and is constant for each trial. More precisely, we expect a binomial distribution under the following conditions;

a) n the number of trials is finite.

b) Trials are independent.

c) θ, the probability of success is constant for each trial, and then $(1-\theta)$ is the probability of failure in any trial.

If y denotes the number of successes in trials satisfying the above conditions, then y is a random variable which can takes the values 0,1,2,…,n; since in n trials we may get no success (all failures), one success, two success,..., or all the n successes.

We are interested in finding the corresponding probabilities of 0,1,2 ,….,n successes the general expression for the probability of r successes are given by;

$$P(r) = P(Y = r) = {}^{n}C_{r}\theta^{r}(1-\theta)^{n-r}; \qquad r = 0,1,\ldots,n$$

ii) Poisson Distribution: Poisson distribution was discovered by the French mathematician and Physicist Simeon Denis Poisson (1781-1840) who published it in 1837. He derived it as a limiting case of binomial distribution. If a dichotomous variable y is such that the constant probability p of success for each trail is very small and the number of trails n is indefinitely large and is $np = \mu$ is finite, the probability of y successes is given by

$$p(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}; \quad y = 0,1,\ldots$$

where λ is known as parameter of Poisson distribution. The mean and variance of Poisson are same i.e. λ the only parameter in Poisson distribution.

iii) Normal distribution: A random variable X is normally distributed with location parameter μ and scale parameter σ if its pdf is given by

$$f(y)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \qquad ;-\infty < y < \infty\ ;\ -\infty < \mu < \infty\ ;\sigma > 0$$

with mean μ and variance σ^2.

The normal distribution curve is bell shape and symmetrical about the line $y=\mu$. The mode and medium of the normal curve lies at the point $y=\mu$. The area under the normal curve within its range $-\infty to \infty$ in always unity i.e. $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{\frac{(y-\mu)^2}{2\sigma^2}\right\}dy=1$. One of the greatest reasons behind the extensive use & application of normal distribution lies in central limit theorem which states: If $y_1, y_2, ..., y_n$ is a random sample of size n from any population with mean μ and variance σ^2. The distribution of sample mean $\bar{y}$ is asymptotically normal with mean μ and variance σ^2/n as $n\to\infty$. Almost all sampling distributions like t, χ^2, F etc., for their large degrees of freedom conform to normal distributions.

CHAPTER – 2

PRIOR DISTRIBUTION AND ESTIMATION TECHNIQUES

2.1 Introduction:

The fundamental part of any Bayesian analysis is the prior distribution. The prior distribution $\mathrm{P}(\theta)$ represents all that is known or assumed about the parameter θ usually the prior information is subjective and is based on a person's own experience and judgment, a statement of one's degree of belief regarding the parameter, design information and personal opinions. The other critical feature of the Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even a bad prior will still not greatly influence the posterior. In a sense, this is an asymptotic property of Bayesian analysis in that all but pathological priors will be overcome by sufficient amounts of data. We can check the impact of the prior by seeing how stable to posterior distribution is to different choices of priors. If the posterior is highly dependent on the prior, then the data (the likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contains significant information.

Prior distribution may be categorical in different ways. One common classification is a dichotomy that separated "proper" and "improper" priors. A prior distribution is proper if it does not depend on the data and the value of integral $\int_{-\infty}^{\infty} P(\theta)d\theta$ or summation $\sum P(\theta)$ is one. If the prior does not depend on the data and the distribution does not integrate or sum to one then we say that the prior is improper. Other classification of prior is either based on properties or on distributional forms as under:

i. Uniform prior:

In a state of ignorance the prior distribution is accepted as being uniform. It appears that great minds like Gauss, Bernoulli and Laplace used the principle in some form or other in their work. It is claimed that Bayes himself used uniform prior in his revolutionary work.

The apparent success with uniform prior subscribed to the senore's idea that perhaps the uniform prior is the final answer. Jeffery's (1961) makes an interesting comment that there is no more need for such an idea than to suggest that an oven which cooked roast beef once cannot cook anything other than roast beef. One should be cautious before invoking the uniform prior theory, for a careless and mechanical use of this principal may lead to contradiction and confusion.

ii. Non informative prior (NIP):

One class of prior distribution is called non-informative prior and as the name suggests, it is prior that contains no information about Θ. Non informative priors are also called priors of ignorance Box and Tiao (1973) provides a thorough discussion of non informative priors for one or more parameters.

Rather than a state of a complete ignorance, the non informative prior refers to the case when relatively little (or very limited) information is available a priori. In other words, a priori information about the parameter is not substantial relative to the information expected to be provided by the sample of empirical data. A prior probability distribution that represents perfect ignorance or indifference would produce a posterior probability distribution that represents what one should need about the parameter Θ on the basis of the evidence (data) Y alone. Such a prior is called "neutral" or non informative priors by Royall (1997). According to Jeffery's (1983), non -informative priors provide a formal way of expressing ignorance of the value of the parameter over the permitted range.

If the prior is non informative, we should assign the same density to each $\theta \in \Omega$, which of course implies that prior $P(\theta)$ is uniform given by $P(\theta = k, \theta \in \Omega)$.

The non informative prior often leads to a class of improper prior, improper in the sense that $\int_{\Omega} P(\theta) d\theta \neq 1$. The derivation of non informative prior is mathematically very closely associated with variance stabilizing transformations (Bartlett, 1937) and Fishers information (fisher, 1922, 1925).

iii. Natural conjugate prior (NCP):

Raiffa and Schlaifer (1961) presented a formal development of conjugate prior distribution, intuitively, a conjugate prior distribution; say $P(\theta)$ for given sampling distribution, say $f(Y|\theta)$ is such that the posterior distribution $P(\theta|Y)$ and the prior $P(\theta)$ are members of the same family of distributions.

Let $\underline{Y} = (y_1, y_2, ..., y_n)$ be a data from some family of distribution $f(Y|\theta)$ which combines basic information. Such a function is known as sufficient statistic. Sufficient statistic exists for a number of standard distributions.

As in frequencies frame work, sufficient statistic plays an important role in Bayesian interference in constructing a family of prior distribution known as natural conjugate prior (NCP) .The family of prior distribution $P(\theta), \theta \in \Omega$ is called a natural conjugate family if the corresponding posterior distribution belongs to the same family as $P(\theta)$.

The below given table has shown the conjugate priors for several common likelihood functions.

Conjugate prior for common likelihood functions

Likelihood	**Conjugate priors**
Binomial	Beta
Multinomial	Dirichlet
Poisson	Gamma
Normal μ Unknown, σ^2 known	Normal
Normal μ Known, σ^2 unknown	Inverse chi-square
Multivariate normal μ unknown, v known	Multivariate normal
Multivariate normal μ known, v unknown	Inverse Wishart

iv. Jeffrey's Invariant Prior (1946, 1961):

In situations where we only have limited data available and we have no expert knowledge available. We should be able in such situations to choose a suitable prior which should obey the invariant property under parameter transformation. The Jeffery prior was designed to solve the invariance under the parameter transformation problem. According to the Jeffery principal the following equation should hold:

$$P(\Phi) = P(\theta)\left|\frac{d\theta}{d\phi}\right| = P(\theta)\,\left|h'(\theta)\right|^{-1}$$

where $\Phi = h(\theta)$ is a one to one parameter transformation. This states that a rule for determining a prior should yield an equivalent transformed. From the above formulation we can derive the general formula of the Jeffery prior, which is given as

$$P(\theta) \propto \sqrt{I(\theta)} \propto \left[-E\left\{\frac{\partial^2 \log L(\theta \mid Y)}{\partial \theta^2}\right\}\right]^{1/2}$$

where $I(\theta)$ is the Fisher information for the parameter θ. When there are multiple parameters I is the Fisher information matrix, the matrix of the expected second partials

$$I(\boldsymbol{\theta}) = E\left\{\frac{\partial^2 \log L(\theta \mid Y)}{\partial\theta_i \partial\theta_j}\right\}$$

In this case, the Jeffery prior becomes

$$P(\boldsymbol{\theta}) \propto \sqrt{\det[I(\boldsymbol{\theta})]}$$

Jeffery prior for the common probability distribution

Probability Distribution	**Jeffery's prior**
Normal μ Unknown, σ^2 known	$P(\mu) =$ constant
Normal μ Known, σ^2 unknown	$P(\sigma) \propto \frac{1}{\sigma}$
Normal μ, σ^2 Both unknown	$P(\mu,\sigma) = P(\sigma)P(\mu \mid \sigma) = P(\sigma)P(\mu)$
Exponential Distribution	$P(\theta) \propto \frac{1}{\theta}$
Binomial Distribution with n independent draws	$P(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$
Weibull (θ, ϕ)	$P(\phi,\theta) = P_1(p)P_2(\theta \mid \phi)$ $P(\phi,\theta) \propto \frac{1}{\phi^\theta}$
Negative Binomial Distribution	$P(\theta) \propto \theta^{-1}(1-\theta)^{-1/2}$
Uniform Distribution i.e. $X \sim U(0,\theta)$	$P(\theta) \propto \frac{1}{\theta}$

v. Maximal information prior (MIP):

Zellner (1977) used the information theoretic approach to define maximal information prior. Let $I_y(\theta) = \int f(y \mid \theta) \log f(y \mid \theta) d\theta$ be a measure of information in the pdf $f(y \mid \theta)$. The prior average information is defined as

$$\bar{I}_y(\theta) = \int I_y(\theta) P(\theta) d\theta$$

where $P(\theta)$ is a prior density of θ and $\int P(\theta) \log P(\theta) d\theta$ measures the information in prior $P(\theta)$.

$$G = \bar{I}_y(\theta) - \int P(\theta)\log P(\theta)d\theta$$

$$G = \int I_y(\theta)P(\theta)d\theta - \int P(\theta)\log P(\theta)d\theta$$

is defined as a measure of gain in information, the maximal information prior is the one that maximizes G for varying $P(\theta)$ subject to the condition $\int P(\theta)d\theta = 1$

vi. Asymptotically locally invariant prior (ALIP):

Hartigan (1964) derived a family of prior densities to represent our ignorance about θ using invariance techniques similar to those suggested by Jeffery's (1946). He named this asymptotically locally invariant (ALI) prior. The ALI priors are easy to derive for exponential family of distributions.

Hartigan (1964) point out that in some instances, the posterior distribution based on the ALI prior may lead to a chi-square having a degree of freedom contrary to the usual rule of assigning the degree of freedom to chi-square.

vii. Dirichlets prior (DP):

Dirichlets prior distribution is

$$P(p_1, p_2, \ldots, p_k) = \frac{\Gamma\theta}{\Gamma\theta_1.\Gamma\theta_2\ldots\Gamma\theta_k} p_1^{\theta_1 - 1}.p_2^{\theta_2 - 1}\ldots p_k^{\theta_k - 1}$$

where $\theta = \sum_{i=1}^{k}\theta_i$, $\sum_{i=1}^{k} p_i = 1, ; 0 < p_i < 1; \theta_i > 0$ is a generalization of the beta –prior.

$$P(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}, 0 < \theta < 1; a, b > 0$$

viii. Haldane's prior (1931):

Halden's prior is given as

$$P(\theta) \propto \theta^{-1}(1-\theta)^{-1}, \quad \theta \in [0,1]$$

which is an improper density. We get if we put $\alpha = \beta = 0$ in Beta prior.

2.2 Loss Function (Lf):

The word "loss" is used in place of "error" and the loss function is used as a measure of the error or loss. Let θ be an unknown parameter of some distribution $f(y \mid \theta)$ and suppose that θ is estimated by some statistics $T(Y) \cong T$. Let $L(\theta, T)$

represents the loss incurred when the true value of the parameter is Θ where Θ is estimated by the statistics T.

Loss function is a measure of the error and presumably would be greater for large error than for small error. We would want the loss to be small or we want the estimate to be close to what it is estimating. Our objective is to select an estimator T= L($y_1, y_2, ..., y_n$) that makes this error or loss small. Loss depends on sample and we cannot hope to make the loss small for every possible sample but can try to make the loss small on the average. Our objective is to select an estimator that makes the average loss (risk) small and ideally select an estimator that has the small risk.

Some Important Loss Functions are as under:

a) Squared-Error Loss Function:

The squared error loss function (SELF) was proposed by Legendre (1805) and Gauss to develop least squares theory. Later, it was used in estimation problems when unbiased estimators of θ were evaluated in terms of the risk function $\mathrm{R}(\theta, \mathrm{T})$ which becomes nothing but the variance of the estimator. It was also observed that SELF is a convex loss function and therefore, restricts the class of estimators by excluding randomized estimator. The SELF is given as

$$L(\theta, T) = (\theta - T)^2 .$$

b) Weighted SELF:

A generalization of squared-error loss, which is of interest, is

$$L(\theta, T) = W(\theta)(\theta - T)^2$$

This loss is called weighted squared-error loss and has the attractive feature of allowing the squared error, $(\theta - T)^2$ to be weighted by a function of $-\Theta$.

c) Quadratic SELF:

Other variant of square error loss is quadratic SELF. If $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$ is a vector to be estimated by $T = (t_1, t_2, ..., t_p)'$, and Q is pxp positive definite matrix, and then $L(\boldsymbol{\theta}, T) = (\boldsymbol{\theta} - T)' Q (\boldsymbol{\theta} - T)$ is called quadratic loss. When Q is diagonal, this reduces to

$$L(\boldsymbol{\theta}, T) = \sum_{i=1}^{k} Q_i \left(\theta_i - t_i\right)^2$$

and is natural extension of squared-error loss to the multiparametric situation.

d) Linear Loss:

When utility function is approximately linear (as is often the case over a reasonable segment of the reward space), the loss function will tend to be linear. Thus of interest is the linear loss

$$L(\theta, T) = C_1(\theta - T), \quad \theta \geq T$$

and $$L(\theta, T) = C_2(T - \theta); \quad \theta < T$$

The constants C_1 and C_2 reflect the effect of over and over estimating θ. By suitably choosing C_1, C_2 any fractile of the posterior distribution will be a Bayes estimator (Box and Tiao, 1973).

If C_1 and C_2 are functions of θ, the above loss function is called weighted linear loss function.

e) Absolute Loss:

$$L(\theta, T) = |\theta - T|$$

is called the absolute loss function. As per De Groot (1970) for such a loss function, Bayesian estimator is the posterior median.

f) Zero –One loss:

$$L(\theta, T) = 0 \qquad \text{iff } |\theta - T| \leq C$$

and

$$L(\theta, T) = 1 \qquad \text{iff } |\theta - T| > C$$

where c is the small positive constant.

As per Raiffa and Schlaifer (1961), Bayes estimator for such a loss function is mode of posterior distribution. The risk function $R(\theta, T)$, associated with the estimator T is defined as the expected value of the loss function. The loss is Zero if the decision is made correct about T and the loss is one if the decision about T is made incorrect.

$$R(\theta, T) = E_y[L(\theta, T)] = \int L(\theta, T) f(y \mid \theta) dy$$

$$= P\big[|\theta - T| > C\big]$$

= P[incorrect decision about T]

2.2.1 Risk Function:

The risk function $R(\theta,T)$ associated with an estimator T is defined as the expected value of the loss function and is given by

$$R(\theta,T)=E_y\left[L(\theta,T)\right]=\int L(\theta,T)\,f(y\,|\,\theta)dy$$

Bayes risk associated with an estimator T is defined as the expected value of the risk function $\mathrm{R}(\theta,\mathrm{T})$ with respect to the prior distribution $p(\theta)$ of θ and is given by

$$R(\theta,T)=E_\theta\left[R(\theta,T)\right]$$

$$=\int R(\theta,T)P(\theta)d\theta$$

$$=\int\left[\int L(\theta,T)\,f(y\,|\,\theta)\right]P(\theta)d\theta$$

Bayesian risk of an estimator is an average risk, which is a real number. Risk can be used as a guide. A good decision would be that minimizes the risk for all values of θ in Ω. For two estimators $T_1=t_1(y_1,y_2,...,y_n)$ and $T_2=t_2(y_1,y_2,...,y_n)$ estimator T_1 is defined to be better estimator than T_2 if

$$Rt_1(\theta)\le Rt_2(\theta),\forall\theta\in\Omega$$

Thus, risk and loss functions are used to assess the goodness of estimators.

2.3 Estimation Techniques:

The word estimator stands for the function, and the word, estimate stands for a value of that function. In estimator we take a random sample from the distribution to elicit some information about unknown parameter θ. That is, we repeat the experiment n independent times, observe the sample,and we try to estimate the value of θ, using the observations $y_1,y_2,...,y_n$. The function of $y_1,y_2,...,y_n$ used to estimate θ ; say the statistic $U(y_1,y_2,...,y_n)$ called an estimator of θ . We want it to be such that the computed estimate $U(y_1,y_2,...,y_n)$ is usually close to θ .

Thus any statistic whose values are used to estimate $\mathrm{r}(\theta)$ where r(.) is some function of the parameter θ, is defined to be an estimator $\mathrm{r}(\theta)$. An estimator is always a statistic which is both a random variable and a function.

2.4 Methods of estimation:

A variety of methods to estimate the unknown parameters have been proposed. The common used methods are:

i) Method of maximum likelihood estimation,

i) Method of minimum variance,

ii) Method of moment,

iii) Method of least square estimation,

iv) Method of minimum chi-square, and

v) Bayesian estimation.

These methods are described are follows:

i) Method of maximum likelihood estimation (MLE):

The most general method of estimation known is the method of maximum likelihood estimators (MLE) which was initially formulated by C.F.Gauss but as a general method of estimation was first introduced by Professor. R. A. Fisher in the early (1920) and later on developed by him in a series of papers. He demonstrated the advantages of this method by showing that it yields sufficient estimators, which are asymptotically MVUES. Thus the essential feature of this method is that we look at the value of the random sample and then choose our estimate of the unknown population parameter, the value of which the probability of obtaining the observed data is maximum. If the observed data sample values are $(y_1, y_2, ..., y_n)$ we can write in the discrete case.

$$P(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n) = f(y_1, y_{2,} ..., y_n)$$

which is just the value of joint probability distribution of the random values $(y_1, y_2, ..., y_n)$ at the sample point $(y_1, y_2, ..., y_n)$ since the sample values has been observed and are therefore fixed numbers, we regard $f(y_1, y_{2,} ..., y_n; \Theta)$ as the value of a function of the parameter Θ, referred to as the likelihood function. A similar definition applies when the random sample comes from a continuous population but in that case $f(y_1, y_2, ..., y_n; \Theta)$ is the value of joint pdf at the sample point $(y_1, y_2, ..., y_n)$ i.e.; the likelihood function at the sample value $(y_1, y_2, ..., y_n)$

$$L = \prod_{i=1}^{n} f(y_i, \Theta) \qquad (2.4.1)$$

Since the principle of maximum likelihood consists in finding an estimator of the parameter which maximizes L for variation in the parameter. Thus if there exists a function $\hat{\theta} = \hat{\theta}(y_1, y_2, ..., y_n)$ of the sample values which maximizes L for variation in Θ, then $\hat{\theta}$ is to be taken as the estimator of Θ. $\hat{\theta}$ is usually called ML estimators. Thus is the solution if and only if

$$\frac{\partial L}{\partial \theta} = 0 and \frac{\partial^2 L}{\partial \theta^2} < 0 \quad (2.4.2)$$

Since L >0, so LogL which shows that L and Log L attains their extreme values at the $\hat{\theta}$. Therefore, the equation becomes

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0 \quad (2.4.3)$$

a form which is more convenient from practical point of view.

ii)Method of minimum variance [minimum variance unbiased estimators (MVUE)]:

If a statistic $T = T(y_1, y_2, ..., y_n)$, based on sample of size n such that:

a) T is unbiased for $r(\theta)$, for all $\Theta \in \Theta$ and

b) It has the smallest variance among the class of all unbiased estimators of $r(\theta)$,

Then T is called the minimum variance unbiased estimator (MVUE) of $r(\theta)$. More precisely, T is MVUE of $r(\theta)$ if

$$E_\theta = r(\theta) \forall \theta \in \Theta$$

and $Var_\theta(T) \geq Var_\theta(T') \forall \theta \in \Theta$ where T' is any other unbiased estimator of $r(\theta)$

Crammer-Rao in equality provides a lower bound $[r'(\theta)]^2 / I(\theta)$, to the variance of an unbiased estimator of $r(\theta)$, where $I(\theta)$ is the information on Θ, supplied by the sample.

An unbiased estimator t of r (Θ) for which Crammer-Rao lower bound is attained is called a minimum variance bound (MVB) estimator.

The method of minimum variance involves estimates which (i) are unbiased and (ii) have minimum variance.

If $L = \prod_{i=1}^{n} f(y_i, \theta)$, is likelihood function of a random sample n observations $(y_1, y_2, ..., y_n)$ from a population with probability function $f(y, \theta)$, then the problem is to find a statistic $t = t(y_1, y_2, ..., y_n)$ such that

$$E(t) = \int_{-\infty}^{\infty} t.L dy = r(\theta) \Rightarrow \int_{-\infty}^{\infty} [t - r(\theta)] L.dy = 0 \qquad (2.4.4)$$

and $$V(t) = \int_{-\infty}^{\infty} [t - E(t)]^2 L.dy = \int_{-\infty}^{\infty} [t - r(\theta)]^2 L.dy \qquad (2.4.5)$$

is minimum, where $\int_{-\infty}^{\infty} dy$ represents the n-fold integration

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dy_1 . dy_2 \ldots dy_n \qquad (2.4.6)$$

Thus, we have to minimize V(t) subject to the condition E(t).

(iii) Method of moments (substitution principle) (MM):

One of the simplest and oldest methods of estimation is the substitution principle. The method of moments was discovered and studied in detail by Karl Pearson. The method of moments is special case when we need to estimate some known function of finite number of unknown moments.

Let $f(y;\theta_1,\theta_2,\ldots,\theta_k)$ be density function of the parent population with k parameters $\theta_1,\theta_2,..,\theta_k$. If μ_r' denotes the rth moment about origin, then

$$\mu_r' = \int_{-\infty}^{\infty} y^r f(y;\theta_1,\theta_2,\ldots,\theta_k), r = 1,2,\ldots.,k \qquad (2.4.7)$$

In general $\mu_1',\mu_2',\ldots,\mu_k'$ will be functions of the parameters $\Theta_1,\Theta_2,\ldots,\Theta_k$. Let $y_i, i = 1,2,3,\ldots,n$ n be a random sample of size n from the given population. The method of moments consists in solving the k-equation (i) for $\Theta_1,\Theta_2,\ldots,\Theta_k$ in terms of $\mu_1',\mu_2',\ldots,\mu_k'$ and then replacing these moments.

$\mu_r'; r = 1,2,3,\ldots,k$ by the sample moments

e.g. $\hat{\theta}_i = \hat{\theta}(\hat{\mu}_1', \hat{\mu}_2', \ldots, \hat{\mu}_k') = \theta_i(m_1', m_2', \ldots, m_k'); i = 1,2,\ldots,k$ (2.4.8)

where m_i is the ith moment about origin in the sample.

Then by the method of moments $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ are the estimators of respectively.

(iv) Method of least square estimation (LSE):

The principle of least square is used to fit a curve of the form

$$y = f(y; a_0, a_2, ..., a_n)$$

where $a_i's$ are unknown parameters, to a set of n sample observations $(x_i, y_j); i = 1,2,3,...,n$ from a bivariate population. It consists in minimizing the sum of squares of residuals viz.,

$$E = \sum_{i=1}^{n} [y_i - f(x_i; a_0, a_1, ..., a_n)]^2$$

Subject to variations in $a_0, a_2, ..., a_n$

The normal equations for estimating $a_0, a_2, ..., a_n$ are given by

$$\frac{\partial E}{\partial a_i} = 0; i = 1,2,...,n \qquad (2.4.9)$$

(v) Methods of minimum chi-square (MC):

If the observations are grouped in 'k' mutually exclusive classes with frequencies $(f_1, f_2, ..., f_k)$ with $\sum_i f_i = n$. Suppose the unknown probabilities of these classes are $p_1, p_2, ..., p_k$ which depends on the parameter $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$. The problem is to estimate $\underline{\theta}$ The expected frequencies of this k -classes will be $np_1, np_2, ..., np_k$. If measure of discrepancy between the set of observed frequencies and the corresponding expected frequencies are provided by χ^2 defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{[f_i - np_i(\theta)]^2}{np_i(\theta)} = \frac{\sum f_i^2}{np_i(\theta)} - n$$

The measure of discrepancy is a function of unknown parameters $\underline{\theta}$.

$$\chi^2 = (\theta_1, ..., \theta_I)$$

The method of minimum χ^2 is to take that estimate of θ_j which minimizes χ^2.

Thus we solve the equation

$$\frac{\partial \chi^2}{\partial \theta_j} = 0; j = 1,2, ..., l$$

$$\frac{\partial}{\partial \theta_j}\left[\sum\left(\frac{f_i^2}{np_i(\theta)}\right) - n\right] = 0 \qquad (2.4.10)$$

Let $\hat{\theta}$ be the solution of equation (I) and satisfy

$$\left[\frac{\partial^2\chi^2}{\partial\theta_j^2}\right]_{\theta=\hat{\theta}} > 0$$

Then $\hat{\theta}$ will be minimum χ^2 estimate of θ .

2.5 Bayesian method of estimation:

Suppose we have a random sample of size n say $y_1, y_2, ..., y_n$ which we regard as independent identically distributed random variables with distribution function $(df)\, F(Y\,|\,\theta)$ and pdf $f(y\,|\,\theta)$ and where θ a labeling parameter, real valued or vector valued as the case may be. Also we assume that we do not know the exact value of parameter θ. There are cases in which one can assume a little more about the unknown parameter $\theta \in \Omega$. Here Ω is the parameter space. We could assume that θ is itself a random variable with distribution function $F(\theta)$ or pdf $P(\theta)$ Now suppose n items are put to test and it is assumed that their recorded life times from a random sample of size n from a population with pdf $f(y\,|\,\theta)$.To be specific we will assume θ to be real valued. We agree to regard θ itself as random variable with a pdf $P(\theta)$. The joint pdf of $P(\theta)$ is given by

$$P(y_1, y_2, ..., y_n \,|\, \theta) = \left\{\prod_{i=1}^{n} f(y_i \,|\, \theta)\right\} = L(y_1, y_2, ..., y_n \,|\, \theta)$$

The marginal pdf of $y_1, y_2, ..., y_n$ is given by

$$p(y_1, y_2, ..., y_n) = \int_{\Omega} p(y_1, y_2, ..., y_n \,|\theta) d\theta$$

and the conditional pdf of θ given data $y_1, y_2, ..., y_n$ is given by

$$P(\theta \,|\, y_1, y_2, ..., y_n) = \frac{p(y_1, y_2, ..., y_n \,|\, \theta)}{p(y_1, y_2, ..., y_n)}$$

$$P(\theta \,|\, y_1, y_2, ..., y_n) = \frac{L(y_1, y_2, ..., y_n \,|\, \theta)\, p(\theta)}{\int_{\Omega} L(y_1, y_2, ..., y_n \,|\, \theta)\, p(\theta) d\theta}$$

Thus, prior to obtaining the data $y_1, y_2, ..., y_n$ the variations in θ where represented by $P(\theta)$, known as prior distribution on θ however, after the data $y_1, y_2, ..., y_n$ has been obtained in the light of the new information, the variation in θ are represented by $P(\theta \,|\, y_1, y_2, ..., y_n)$ the posterior distribution of θ. The uncertainty about the parameter θ Prior to experiment is represented by prior pdf $P(\Theta)$ and the same after the experiment is represented by posterior pdf $P(\Theta | y_1, y_2, ..., y_n)$. This process is the straight forward application of the Bayes theorem. Once the posterior distribution has been obtained it becomes the main object of study.

2.6 Large sample approximations:

In many areas of application, simple models suffice for most practical purposes but there are occasions when the complexity of the scientific questions at issue and the data available to answer them warrant the development of more sophisticated models, which depart from standard forms. For such models, approximations to the posterior distribution of model parameters are useful in their own right and as a starting point for more exact methods. The approximations that we describe are relatively easy to compute, understand and can provide valuable information about the fit of the model. Some important methods of approximation are given below

a) Normal approximation to posterior distribution:

The numerical implementation of a Bayesian procedure is not always straight forward since the involved posterior distribution is complicate functions. One of the important steps in simplifying the computations is to investigate the large sample behavior of the posterior distribution and its characteristics. The basic result of the large sample Bayesian inference is that the posterior distribution of the parameter approaches a normal distribution. Relatively little has been written on the practical implications of asymptotic theory for Bayesian analysis. The overview by Edwards, Lindeman, and Savage (1963) remains one of the best and includes a detailed discussion of the principle of 'stable estimation' or when prior information can be satisfactorily approximated by a uniform density function. Some good sources on the topic from the Bayesian point of view include Lindley (1958), Pratt (1965), and Berger and Wolpert (1984). An example of the use of the normal approximation with small samples is provided by Rubin and Schenker (1987), who approximated the posterior distribution of the logit of the binomial parameter in real application and evaluate the frequentists operating characteristics of their procedure. Clogg *et al.,* (1991) provide additional discussion of this approach in a more complicated setting. Sequential monitoring and analysis of clinical trials in medical research is an important area of practical application that has been dominated by frequentists thinking but has recently seen considerable discussion of the merits of a Bayesian approach; a recent review is provided by Freedman, Spiegel halter and Parmer (1994), Khan, A.A (1997) and Khan *et al.,* (1996).

If the posterior distribution $P(\theta \mid y)$ is unimodal and roughly symmetric, it is convenient to approximate it by a normal distribution centered at the mode; that is logarithm of the posterior is approximated by a quadratic function, yielding the approximation

$$P(\theta \mid y) \sim N\left(\hat{\theta}, \left[I(\hat{\theta})\right]^{-1}\right)$$

where $I(\hat{\theta}) = -\dfrac{\partial^2 \log P(\theta \mid y)}{\partial \theta^2}$ (2.6.1)

if the mode, $\hat{\theta}$ is in the interior parameter space, then $I(\theta)$ is positive; if $\hat{\theta}$ is a vector parameter, then $I(\theta)$ is a matrix.

b)Lindley's Approximation (1980):

Many times the integrals appearing in Bayes estimation cannot be expressed in a closed form when the chosen prior distribution is conjugate priors. In particular, we come across evaluation of posterior expected value of $U(\theta)$ which involves ratio of the integrals

$$\int_{\Omega} U(\theta)L(\theta \mid Y)P(\theta)d\theta \text{ and } \int_{\Omega} L(\theta \mid Y)P(\theta)d\theta .$$

Lindley (1980) considered evaluation of the ratio of the integrals $\int_{\Omega} U(\theta)L(\theta \mid Y)P(\theta)d\theta$ and $\int_{\Omega} L(\theta \mid Y)P(\theta)d\theta$ which is nothing but $E[U(\theta) \mid Y]$.Let us consider the case of a scalar parameter θ of the distribution having pdf (pmf) $f(Y \mid \theta)$. Suppose the likelihood function has a unique maximum $\hat{\theta}$ maximum likelihood estimate of Θ .

We have, $$E(U(\theta) \mid Y) = \frac{\int_{\Omega} U(\theta)L(\theta \mid Y)P(\theta)d\theta}{\int_{\Omega} L(\theta \mid Y)P(\theta)d\theta}$$

$$U_i(\theta) = \left(\frac{\partial}{\partial \theta}\right)^i U(\theta); L(\theta) = \log L(\theta \mid Y)$$

For $\sigma^2 = \left(\frac{\partial^2}{\partial\theta^2} L(\theta)\right)^{-1}$; $L_i(\theta) = \frac{\partial^i}{\partial\theta^i} L(\theta)$

$\rho(\theta) = \log P(\theta)$ and $\rho_1(\theta) = \frac{\partial}{\partial\theta}[\rho(\theta)]$

Then, Lindley's approximation, for large n of $E(U(\theta) | Y)$ is given by

$$E(U(\theta) | Y) \cong [U(\theta) + \frac{1}{2}(U_2(\theta) + 2U_1(\theta)\rho_1(\theta)\sigma^2 + \frac{\sigma^4}{2} L_3(\theta)U_1(\theta)]_{\theta=\hat{\theta}}$$

approximation is of $O\left(\frac{1}{n^2}\right)$

c) Laplace Approximation:

A simple and remarkable method of asymptotic expansion of integrals generally attributed to Laplace (Laplace, 1986, 1774, Stigler, 1986) is widely used in applied mathematics. This method has been applied by many authors (Lindley, 1961, 1980; Mostller and Wallace, 1964; Johnson, 1970; DiCiccio, 1986; Hartigan, 1965; Khan et al., 1996; and Tierney and Kadane, 1986 and Yoichi Miyata, 2004) to find approximations to the ratios of integrals of the interest, especially in Bayesian analysis. If we approximate the integrals involved in the posterior density using approximation

$$P(\theta | Y) = (2\pi)^{\frac{-k}{2}} \left|I(\hat{\theta})\right|^{\frac{1}{2}} \exp\left[\log P(\hat{\theta} | y)\right] \left(1 + O(n^{-1})\right) \quad (2.6.1)$$

where $\left|I(\hat{\theta})\right|$ stands for determinant of $I(\hat{\theta})$ then posterior density can be approximated with error of order $O(n^{-1})$ i.e.

$$P(\theta | Y) = (2\pi)^{\frac{-k}{2}} \left|I(\hat{\theta})\right|^{\frac{1}{2}} \exp\left[\log P(\theta | y) - \log P(\hat{\theta} | y)\right] \left(1 + O(n^{-1})\right) \quad (2.6.2)$$

approximation (2.6.2) is the well known Laplace's approximation of integrals (e.g., Tierney and Kadane, 1986). Laplace's approximation (2.6.2) of posterior density can be compared with normal approximation which has error of order $O(n^{-\frac{1}{2}})$. Perhaps more importantly, Laplace's approximation is of order $O(n^{-1})$ uniformly on any neighborhood of the mode. This means that it should provide a good approximation in the tails of distribution also (e.g., Tierney and Kadane, 1986; Tierney, Kass and Kadane, 1989a; and Wong and Li, 1992).

CHAPTER – 3

BAYESIAN ESTIMATION FOR BINOMIAL DISTRIBUTION

3.1 Introduction:

Binomial distribution is also known as the Bernoulli distribution after the Swiss mathematician James Bernoulli (1654-1705) who discovered it in 1700 and was first published in 1713, eight years after his death. This distribution can be used under the following conditions:

v) The random experiment is performed repeatedly a finite and fixed of times. In other words n, the number of trials is finite and fixed.

vi) The outcome of the random experiment (trial) results in the dichotomous classification of events. In other words, the outcome of the trial may be classified into two mutually disjoint categories called success (the occurrence of the event) and failure (the non-occurrence of event).

vii) All the trials are independent, i.e. the result of any trial, is not affected in any way by the preceding trials and does not affect the result of succeeding trials.

viii) The probability of success (happening of event) in any trial is θ and is constant for each trial. $1-\theta$ Is then termed as the probability of failure (non-occurrence of the event and is constant for each trial. More precisely, we expect a binomial distribution under the following conditions:

  d) n the number of trials is finite.

  e) Trials are independent.

  f) θ, the probability of success is constant for each trial, and then $1-\theta$ is the probability of failure in any trial.

If y denotes the number of successes in trials satisfying the above conditions, then y is a random variable which can takes the values 0,1,2,…,n; since in n trials we may get no success (all failures), one success, two success,…., or all the n successes.

We are interested in finding the corresponding probabilities of 0,1,...,n successes the general expression for the probability of y successes are given by:

$$f(Y \mid \theta) = P(Y = y) = {}^{n}C_{y}\theta^{y}(1-\theta)^{n-y};\ y = 0,1,\ldots,n \tag{3.1}$$

Maximum Likelihood Estimation of Binomial Distribution:

Let $y_1, y_2, \ldots, y_n$ be a random sample of size n having the probability mass function given in (3.1) we have

$$f(Y \mid \theta) = {}^{n}C_{y}\theta^{y}(1-\theta)^{n-y};\ y = 0,1,2,\ldots,n$$

Then the Likelihood function is given by

$$L(\theta | Y) = {}^nC_y \theta^y (1-\theta)^{n-y}$$

Applying log on both sides

$$\log L(\theta | Y) = constant + y \log\theta + (n-y)\log(1-\theta)$$

The Mle of θ is the solution of equation

$$\frac{\partial}{\partial\theta}\log L(\theta | Y) = 0$$

$$\hat{\theta} = \frac{y}{n}$$

3.2 Bayes Estimation of Binomial Distribution under Different Types of Priors:

i) Binomial distribution under conjugate prior:

Let $y_1, y_2, ..., y_n$ be a random sample of size n having probability mass function as $f(Y | \theta) = P(Y = y) = {}^nC_y \theta^y (1-\theta)^{n-y}$; $y = 0,1,2,....,n$

Then the likelihood function $L(\theta | Y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$ (3.2.1)

The conjugate prior for θ is

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}; \theta \in [0,1]; \alpha, \beta > 0$$

where α, and β are hyper parameters.

Using Bayes theorem, the posterior distribution of θ is given by

$$P(\theta | Y) \propto l(\theta | y) p(\theta)$$

$$P(\theta | Y) = k\ l(\theta | y) p(\theta)$$

$$= \frac{k\binom{n}{y}\theta^y (1-\theta)^{n-y} \theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$\Rightarrow P(\theta | Y) = \frac{1}{B(\alpha+y, \beta+n-y)} \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$$

The prior and posterior distribution belongs to the family of Beta distribution.

Observe that

The maximum of $l(\theta | y)$ is $\hat{\theta} = \frac{y}{n}$ and that of $P(\theta)$ is given by

$$\frac{\partial}{\partial\theta}\log P(\theta) = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}\left[\log\left(\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)\right] = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}[(\alpha-1)\log\theta+(\beta-1)\log(1-\theta)+\text{constant}]=0$$

$$\Rightarrow \frac{\partial}{\partial\theta}[(\alpha-1)\log\theta+(\beta-1)\log(1-\theta)+\text{constant}]=0$$

$$\Rightarrow \frac{(\alpha-1)}{\theta}-\frac{(\beta-1)}{1-\theta}=0$$

$$\Rightarrow \theta=\frac{(\alpha-1)}{(\alpha+\beta+2)} \quad \text{in } (0,1) \text{ if } \alpha,\beta>1$$

The posterior distribution $P(\theta \mid Y)$ is maximized at

$$\frac{\partial}{\partial\theta}\log P(\theta \mid Y)=0$$

$$\Rightarrow \frac{\partial}{\partial\theta}[(\alpha+y-1)\log\theta+(\beta+n-y-1)\log(1-\theta)+\text{constant}]=0$$

$$\Rightarrow \frac{(\alpha+y-1)}{\theta}-\frac{(\beta+n-y-1)}{1-\theta}=0$$

$$\Rightarrow \theta=\frac{(\alpha+y-1)}{(\alpha+\beta+n-2)}$$

$$\Rightarrow \theta=a\left(\frac{y}{n}\right)+\frac{(1-a)(\alpha-1)}{(\alpha+\beta-2)}$$

where $a=\frac{n}{(\alpha+\beta+n-2)}$

Since $a\in(0,1); P(\theta \mid y)$ synthesizes and compromises by favoring values between the maximum of $p(\theta)$ and that of $l(\theta \mid y)$.

The posterior mean of θ is given by

$$\hat{\theta}_B=E(\theta \mid y)=\int_0^1 \theta\, P(\theta \mid y)d\theta=\int_0^1 \theta\frac{1}{B(\alpha+y,\beta+n-y)}\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}d\theta$$

$$\Rightarrow \hat{\theta}_B=\frac{1}{B(\alpha+y,\beta+n-y)}\int_0^1 \theta^{(\alpha+y+)-1}(1-\theta)^{\beta+n-y-1}d\theta$$

$$\Rightarrow \hat{\theta}_B=\frac{B(\alpha+y+1,\beta+n-y)}{B(\alpha+y,\beta+n-y)}$$

where the Beta function is defined as $B(a,b)=\frac{\Gamma a\,\Gamma b}{\Gamma(a+b)}$

Hence $\hat{\theta}_B=\frac{\Gamma(\alpha+y+1)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta+n+1)}\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)}$

$$\Rightarrow \hat{\theta}_B = \frac{\alpha + y}{\alpha + \beta + n} = \frac{n\left(\frac{y}{n}\right) + (\alpha + \beta)\left(\frac{\alpha}{\alpha + \beta}\right)}{(\alpha + \beta) + n}$$

Also, $\hat{\theta}_B \to \hat{\theta}_{mle}$ *as* $n \to \infty$ for fixed $k = \alpha + \beta$,

$\hat{\theta}_B \to$ *prior mean* $= \dfrac{\alpha}{\alpha + \beta}$ *as* $k \to \infty$ for fixed n.

The variance of the prior distribution $B(\alpha, \beta)$ is given by

$$\operatorname{var}(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} = \frac{\left(\frac{\alpha}{\alpha + \beta}\right)\left(\frac{\beta}{\alpha + \beta}\right)}{(\alpha + \beta + 1)}$$

The form of $\hat{\theta}_B$ and the $\operatorname{var}(\theta)$ of prior distribution suggests that $k = \alpha + \beta$ can be interpreted as the prior sample size. Alternatively it can as be interpreted as since $\hat{\theta}_B$ the m.l.e for data obtained by supplementing the real data (y successes in n trials) by "fictitious data" consisting of α successes in $k = \alpha + \beta$ trials. The quantities α and k need not be integers. Here k plays the role of the prior sample size.

The mode of the $P(\theta \mid y)$ can be regarded as Bayesian m.l.e. Consider

$$\log P(\theta \mid y) \propto (\alpha - 1 + y)\log\theta + (\beta - 1 + n - y)\log(1 - \theta)$$

If $\min(\alpha + y, \beta + n - y) > 1$, then posterior mode is

$$\hat{\theta}_m = \frac{\alpha - 1 + y}{k - 2 + n} = \frac{n}{k - 2 + n}\hat{\theta}_{mle} + \frac{k - 2}{k - 2 + n}\left(\frac{\alpha - 1}{k - 2}\right)$$

is a convex combination of $\hat{\theta}_{mle}$ and expression involving α and β. If $\min(\alpha, \beta) > 1$, $\hat{\theta}_m$ is formally the Bayes estimate corresponding to the prior Beta $(\alpha - 1, \beta - 1)$. This implies $\hat{\theta}_m$ can be regarded as a Bayes estimate under SELF biased on loss certain prior information than the $\hat{\theta}_B$ w.r.t Beta(α, β). Since $\hat{\theta}_m$ has prior sample size $\alpha + \beta - 2$ rather than $\alpha + \beta$

ii) Binomial Distribution under Jeffery's prior:

We have for $Y \sim \operatorname{bin}(n, \theta)$

$$l(\theta \mid Y) = \binom{n}{y}\theta^y (1 - \theta)^{n - y}$$

Then $\log(\theta \mid y) = c + y\log\theta + (n - y)\log(1 - \theta)$

Thus, we have $\dfrac{\partial}{\partial\theta}[\log(\theta \mid y)] = \dfrac{y}{\theta} - \dfrac{n - y}{1 - \theta}$

And likewise $\frac{\partial^2}{\partial\theta^2}[\log(\theta \mid y)] = -\left(\frac{y}{\theta^2} + \left(\frac{n-y}{(1-\theta)^2}\right)\right)$

Since $E(y) = n\theta$, we have

$$-E_y\left[\frac{\partial^2}{\partial\theta^2}[\log(\theta \mid y)]\right] = \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = n\theta^{-1}(1-\theta)^{-1}$$

Hence the Jeffery's prior becomes

$$P(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}}, \quad \theta \in [0,1]$$

This is Beta distribution i.e. $\theta \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$

Thus the posterior distribution by Bayes theorem is

$$P(\theta \mid y) \propto P(\theta) l(\theta \mid y)$$

$$\Rightarrow P(\theta \mid y) \propto \theta^{\frac{-1}{2}} (1-\theta)^{\frac{-1}{2}} \theta^{y} (1-\theta)^{n-y}$$

$$\Rightarrow P(\theta \mid y) \propto \theta^{y+\frac{1}{2}-1} (1-\theta)^{n-y+\frac{1}{2}-1}$$

$$\Rightarrow P(\theta \mid y) = k\theta^{y+\frac{1}{2}-1} (1-\theta)^{n-y+\frac{1}{2}-1}$$

where $k^{-1} = \int_0^1 \theta^{y+\frac{1}{2}-1} (1-\theta)^{n-y+\frac{1}{2}-1} d\theta$

$$k^{-1} = B\left(y + \frac{1}{2}, n - y + \frac{1}{2}\right)$$

Hence the posterior distribution is

$$P(\theta \mid y) = \frac{\theta^{y+\frac{1}{2}-1} (1-\theta)^{n-y+\frac{1}{2}-1}}{B\left(y + \frac{1}{2}, n - y + \frac{1}{2}\right)}$$

Hence the Bayes estimate of θ under Self is given by

$$\hat{\theta}_B = \int_0^1 \theta\, p(\theta \mid y) d\theta$$

$$\hat{\theta}_B = \int_0^1 \theta \frac{\theta^{y+\frac{1}{2}-1} (1-\theta)^{n-y+\frac{1}{2}-1}}{B\left(y + \frac{1}{2}, n - y + \frac{1}{2}\right)} d\theta$$

$$\hat{\theta}_B = \frac{1}{B\left(y + \frac{1}{2}, n - y + \frac{1}{2}\right)} \int_0^1 \theta^{y+\frac{1}{2}+1-1} (1-\theta)^{n-y+\frac{1}{2}-1} d\theta$$

$$\Rightarrow \hat{\theta}_B = \frac{B\left(y+\frac{1}{2}+1, n-y+\frac{1}{2}\right)}{B\left(y+\frac{1}{2}, n-y+\frac{1}{2}\right)}$$

$$\Rightarrow \hat{\theta}_B = \frac{y+\frac{1}{2}}{n+1}$$

which is the Bayes estimate of θ.

iii) Binomial Distribution under Asymptotically Locally Invariant Prior (ALIP):

We have for $Y \sim B(n,\theta)$

$$f(Y\mid\theta)=\binom{n}{y}\theta^{y}(1-\theta)^{n-y},\ y=0,1,2,\ldots,n$$

Since $\log f(Y\mid\theta)=y\log\theta+(n-y)\log(1-\theta)+constant$

$$\log f(Y\mid\theta)=y\log\theta+(n-y)\log(1-\theta)+\text{constant}$$

$$l_1=\frac{\partial}{\partial\theta}\log f(y\mid\theta)=\frac{y}{\theta}-\frac{(n-y)}{(1-\theta)}=\frac{y-n\theta}{\theta(1-\theta)}$$

$$\Rightarrow l_2=\frac{\partial^2}{\partial\theta^2}\log f(y\mid\theta)=-\frac{y}{\theta^2}-\frac{(n-y)}{(1-\theta)^2}$$

we have $E(l_1)=\frac{n\theta}{\theta}-\frac{n-n\theta}{1-\theta}=0$ and $-E(l_2)=\frac{n}{\theta(1-\theta)}$

Also $E(l_1^2+l_2)=\frac{n\theta(1-\theta)}{\theta^2(1-\theta)^2}-\left[\frac{n\theta}{\theta^2}-\frac{n(1-\theta)}{(1-\theta)^2}\right]=0$

Now $E(l_1.l_2)=\frac{-1}{\theta^3(1-\theta)^3}E\left[(x-n\theta)\left\{x(1-\theta-\theta)+n\theta^2\right\}\right]=\frac{-n(1-\theta-\theta)}{\theta^2(1-\theta)^2}$

We have $\frac{\partial}{\partial\theta}\log p(\theta)=-\left[\frac{E(l_1l_2)}{E(l_2)}\right]=-\left[\frac{n(1-\theta-\theta)}{\theta^2(1-\theta)^2}\frac{\theta(1-\theta)}{n}\right]=\frac{1}{1-\theta}-\frac{1}{\theta}$

$$\log p(\theta)\propto-\log(\theta(1-\theta))$$

$$\Rightarrow p(\theta)=\theta^{-1}(1-\theta)^{-1},\ \theta\in[0,1]$$

This is Hartigan’s prior suggested by Haldane(1931).

Then by Bayes theorem posterior distribution is given by

$$p(\theta\mid y)\propto p(\theta)l(\theta\mid y)$$

$$\Rightarrow p(\theta\mid y)\propto\theta^{-1}(1-\theta)^{-1}\theta^{y}(1-\theta)^{n-y}$$

$$\Rightarrow p(\theta\mid y)\propto\theta^{y-1}(1-\theta)^{n-y-1}$$

$$\Rightarrow p(\theta\mid y)=k\theta^{y-1}(1-\theta)^{n-y-1}$$

where $k^{-1} = \int_0^1 \theta^{y-1}(1-\theta)^{n-y-1} d\theta = B(y, n-y)$

Thus the posterior is $p(\theta \mid y) = \frac{1}{B(y, n-y)} \theta^{y-1}(1-\theta)^{n-y-1}$

The Bayes estimate of θ under SELF is as under

$$\hat{\theta}_B = \int_0^1 \theta \, p(\theta \mid y) d\theta$$

$$\Rightarrow \hat{\theta}_B = \int_0^1 \theta \frac{\theta^{y-1}(1-\theta)^{n-y-1}}{B(y, n-y)} d\theta$$

$$\Rightarrow \hat{\theta}_B = \frac{1}{B(y, n-y)} \int_0^1 \theta^{y+1-1}(1-\theta)^{n-y-1} d\theta$$

$$\Rightarrow \hat{\theta}_B = \frac{B(y+1, n-y)}{B(y, n-y)} = \frac{\Gamma(y+1)\Gamma(n-y)\Gamma n}{\Gamma(n+1)\Gamma y \Gamma(n-y)}$$

$$\Rightarrow \hat{\theta}_B = \frac{y}{n}$$

This is the Bayes estimate under (ALIP).

3.3 Improper Marginal posteriors:

Let $Y \sim B(n, \theta)$ where both n and θ are unknown. Suppose n and θ have independent uniform prior distribution

$$f(n, \theta) = f(n) f(\theta)$$

with $f(\theta) = 1$ for $\theta \in [0,1]$ and $f(n) \propto 1$ improper uniform distribution on the positive integers. Then

$$f(n, \theta \mid y) \propto \binom{n}{y} \theta^y (1-\theta)^{n-y} \quad for \; n \geq y$$

Marginal posterior distribution of n is

$$f(n \mid y) = \int_0^1 f(n, \theta \mid y) d\theta \propto \binom{n}{y} \int \theta^y (1-\theta)^{n-y} d\theta$$

$$= \binom{n}{y} B(y+1, n-y+1) = \frac{1}{n+1}, n \in N$$

Since $\sum_{n=x}^{\infty} \frac{1}{n+1} = \infty, f(n \mid y)$ is an improper distribution.

The marginal posterior distribution of θ is

$$f(\theta \mid y) \propto \sum_{n=x}^{\infty} \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} = \sum_{m=0}^{\infty} \binom{m+x-1-1}{m} \theta^{x+1} \frac{(1-\theta)^m}{\theta}$$

Since $\int_0^1 f(\theta \mid y) d\theta \propto \int_0^1 \frac{1}{\theta} d\theta = \infty$

Thus marginal posterior distribution of θ and x are both improper. Thus one has to be cautious in using non-informative prior for more than one parameter case.

3.4 Predictive density:

Let $X \sim B(m,\theta)$ be a future number of successes independent of already observed $Y \sim B(n,\theta)$. Then

$$f(x \mid y) = \int_0^1 f(x,\theta \mid y) d\theta$$

$$\Rightarrow f(x \mid y) = \int_0^1 f(x \mid \theta, y) f(\theta \mid y) d\theta$$

$$\Rightarrow f(x \mid y) = \int_0^1 f(\theta \mid y) f(x \mid \theta) d\theta$$

where x and y have independent conditional distributions and fixed θ.

Therefore $f(x \mid y) = \int_0^1 \binom{m}{x} \frac{\theta^x (1-\theta)^{m-x} \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1}}{B(\alpha+y, \beta+n-y)} d\theta$

$$\Rightarrow f(x \mid y) = \binom{m}{x} \int_0^1 \frac{\theta^{\alpha+y+x-1} (1-\theta)^{\beta+n+m-y-x-1}}{B(\alpha+y, \beta+n-y)} d\theta$$

$$\Rightarrow f(x \mid y) = \binom{m}{x} \frac{B(\alpha+y+x, \beta+n+m-y-x)}{B(\alpha+y, \beta+n-y)}$$

$$\Rightarrow f(x \mid y) = \binom{m}{x} \frac{B(\alpha^*+x, \beta^*+m-x)}{B(\alpha^*, \beta^*)}$$

where $\alpha^* = \alpha + y$ and $\beta^* = \beta + n - y$.

It is known as Beta-binomial distribution with

$$E(X \mid y) = E_{\theta|y} E(X \mid \theta) = E_{\theta|y} E(m\theta) = m \frac{\alpha^*}{\alpha^* + \beta^*}$$

$$E(X \mid y) = \frac{m\left[\left(n - \frac{\bar{y}}{n}\right) + (\alpha+\beta)\frac{\alpha}{(\alpha+\beta)}\right]}{\alpha+\beta+n}$$

and $\operatorname{var}(X \mid y) = E_{\theta|y} \operatorname{var}(X \mid \theta) + \operatorname{var}_{\theta|y}(X \mid \theta)$

$$= E_{\theta|y}(m\Theta(1-\Theta)) + \operatorname{var}_{\theta|y}(m\Theta)$$

$$= m\gamma^*(1-\gamma^*) - m \operatorname{var}(\theta \mid y) + m^2 \operatorname{var}(\theta \mid y)$$

$$\text{var}(X \mid y) = \frac{m\gamma^*(1-\gamma^*)(r^*+m)}{r^*+1}$$

where $\gamma^* = \frac{\alpha^*}{\alpha^*+\beta^*}$, $\text{r}^* = \text{r} + \text{n} = \alpha + \beta + \text{n}.(\text{posteriorsample size})$

we now will be assessing a prior distribution

Let $\text{P}(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\text{B}(\alpha,\beta)}; 0 < \theta < 1;\ 0 < \alpha, \beta < \infty$

with $\gamma = E(\theta) = \frac{\alpha}{\alpha+\beta}, \Phi = \text{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta)} = \frac{\gamma(1-\gamma)}{r+1}$

where $\text{r} = \alpha + \beta$ is known prior sample size.

Since $r = \Phi^{-1} y(1-y) - 1, \alpha = r\gamma$ *and* $\beta = r(1-\gamma)$.

Hence for $\gamma = 0.02$ *and* $\Phi^{1/2} = 0.015$. We have $\text{r} = 86.11, \alpha = 1.72$ and $\beta = 84.39$. This prior information is roughly equivalent to the information. That if you observed a random sample of about 86 individuals out of the population, you would believe that out of the 86 between 1 and 2 were successes. The prior sample size measures the strength of the prior information.

3.5 Estimation of sample size:

Let us suppose $\text{Y} \sim \text{B}(\text{n}, \theta)$

Then Bayes estimation of $P[Y = k(\theta)] = \binom{n}{k}\theta^k(1-\theta)^{n-k}, k = 0,1,2,\ldots,n$

under SELF with Beta(α,β) prior is $\text{P}(\theta \mid \text{y}) = \text{Beta}(\alpha + \text{y}, \beta + \text{n} - \text{y})$

and $E[P[Y = k \mid \theta]] = \int_0^1 \binom{n}{k} \frac{\theta^{\alpha+y+k-1}(1-\theta)^{\beta+n-y+n-k-1}}{B(\alpha+y, \beta+n-y)} d\theta$

$$E[P[Y = k \mid \theta]] = \binom{n}{k} \frac{B(\alpha+y+k, \beta+2n-y-k)}{B(\alpha+y, \beta+n-y)}$$

Let $\text{y} = \sum_{\text{i}=1}^{\text{n}} \text{y}_\text{i}$ then $\text{E}(\theta \mid \text{y}) = \frac{\left(\alpha + \sum_{\text{i}=1}^{\text{n}} \text{y}_\text{i}\right)}{\alpha+\beta+\text{n}}$ and $\text{var}(\theta \mid \text{y}) = \frac{\left(\alpha + \sum_{\text{i}=1}^{\text{n}} \text{y}_\text{i}\right)\left(\beta + \text{n} - \sum_{\text{i}=1}^{\text{n}} \text{y}_\text{i}\right)}{(\alpha+\beta+\text{n})^2(\alpha+\beta+\text{n}+1)}$

Then $\text{r}(\text{f}) = $ Bayes risk of $\text{f}(\theta)$.

$$\Rightarrow r(f) = E^{m(x)} \text{var}(\theta \mid y)$$

$$\Rightarrow r(f) = \frac{E^{m(x)}\left[\alpha(\beta+n) + (\beta+n-\alpha)y - y^2\right]}{k}$$

where $k=(\alpha+\beta+n)^2(\alpha+\beta+n+1)$

Since $E(y)=\sum_{i=1}^{n}E(y_i)=\sum_{i=1}^{n}E\,E(y_i\mid\theta)=E(n\theta)=\frac{n\alpha}{\alpha+\beta}$

$$\Rightarrow E\left(y^2\right)=\sum_{i=1}^{n}E\,E\left(y^2\mid\theta\right)=E\left(n\theta(1-\theta)+n^2\theta^2\right)$$

$$\Rightarrow E\left(y^2\right)=\frac{n\alpha(n\alpha+\beta+n)}{(\alpha+\beta+1)(\alpha+\beta)}$$

So $r(f)=\frac{\alpha\beta\left[(\alpha+\beta)^2+(\alpha+\beta)+2n(\alpha+\beta)+n(n+1)\right]}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)(\alpha+\beta)(\alpha+\beta+1)}$

If c is the cost per observation, then total risk cost is

$$T(n)=r(f)+nc,\frac{\partial T(n)}{\partial n}=0 \text{ gives } \hat{n}=\left[\frac{\alpha\beta}{c(\alpha+\beta)(\alpha+\beta+1)}\right]^{1/2}-(\alpha+\beta).$$

3.6 Estimation of n:

Suppose $y_1,y_2,\ldots,y_r$ are r independent $B(n,\theta)$ random variables. Given the observations $y_1,y_2,\ldots,y_r$ the objective is to estimate n. Suppose for example, that the Apex Appliance Company wishes to estimate the number of a certain type of appliance in use in a certain service area. Suppose further that the company believes that the weekly total of defective appliances sent in for repair (irrespective of age) arises with a binomial probability θ about whose value they have some prior knowledge .Then a count y of the number of defective appliances received during a routine week could be used to caste light on the population size n. In general, then if we have a characteristic with binomial behavior and only the success (or failures) become apparent, we can use these alone to provide information on the population size (Draper and Smith 1971). Adopting their notations, the likelihood can be written as:

$$L(n,\theta\mid y)\propto\theta^t(1-\theta)^{rn-t}\prod_{i=1}^{r}\frac{n!}{y_i!(n-y_i)!} \tag{3.6.1}$$

where $y=(y_1,y_2,..,y_r)'$ is a column vector of positive integers and $t=\sum_{i=1}^{r}y_i$ is the total number of success in the r observation. We now discus the cases θ known and θ unknown separately.

θ *is known :* when θ is known, let $p(n)$ denote the prior distribution of n. Without further knowledge of n, the discrete uniform distribution provides a reasonable form for $p(n)$

$$p(n) = \frac{1}{N}; 1 \le n \le N$$

$$= 0, elsewhere \tag{3.6.2}$$

where N is a large predetermined integer (for example if n were the number of local people of a certain type involved in a certain binomial process, N could be the local population) the posterior distribution for n is then

$$P(n \mid y, \theta) \propto (1-\theta)^{rn} p(n) \prod_{i=1}^{r} \frac{n!}{(n - x_i)!} \tag{3.6.3}$$

The domain of $\mathrm{P}(\mathrm{n} \mid \mathrm{y}, \theta)$ is a set of n such that $n = y_{max}, y_{max} + 1, y_{max} + 2, \ldots, N$ where $y_{max} = \max(y_1, y_2, \ldots, y_r)$.

The mode of the posterior distribution $\mathrm{P}(\mathrm{n} \mid \mathrm{y}, \theta)$ given in expression (3.6.3) denoted by $\hat{n}$ provides an estimate of n. $\hat{n}$ therefore is the integer satisfying the following inequalities:

$$P(\hat{n} - 1 \mid y, \theta) \le P(\hat{n} \mid y, \theta) \text{ and } P(\hat{n} + 1 \mid y, \theta) \le P(\hat{n} \mid y, \theta)$$

or alternatively $\hat{n}$ is the solution of the simultaneous inequalities:

$$\prod_{i=1}^{r} (\hat{n} - y_i) \le [\hat{n}(1-\theta)]^r \quad and \quad \prod_{i=1}^{r} (\hat{n} + 1 - y_i) > [(\hat{n}+1)(1-\theta)]^r$$

which is identical to the maximum likelihood solution (Feldman and Fox, 1968), as expected.

In addition to providing an estimate for n. the posterior distribution could also caste some light on the precession of the estimate. A closed form of the estimator for n may not seem feasible. But a numerical solution can be obtained by using the following recurrence formula. For $n = y_{max} + j, j = 0,1,2 \ldots, (n - y_{max})$

$$P(n \mid y, \theta) = P(y_{max} + j \mid y, \theta) = c\ k_j, \tag{3.6.4}$$

$$where \quad k_j = 1 \quad if \quad j = 0$$

$$= k_{j-i} (1-\theta)^r \frac{(y_{max} + j)^r}{\prod_{i=1}^{r} (y_{max} + j - y_i)}, \; otherwise \tag{3.6.5}$$

Thus, the normalizing constant c in expression (3.6.4) is the reciprocal of the sum of the $k_j's$ i.e. $c = \frac{1}{\sum_{j=0}^{N-y_{max}} k_j}$ (3.6.6)

as for as the point estimation is concerned, an estimate for n can be obtained irrespective of the predetermined integer N. If a confidence interval with a specified confidence coefficient α is desired, then the value of N is needed. A 100α-percent confidence interval for n is given by

$$[y_{max} + l, y_{max} + \mu] \qquad (3.6.7)$$

where l *and* μ are integers such that

$$\sum_{j=1}^{\mu} P(y_{max} + j \mid y, \theta) = \alpha \qquad (3.6.8)$$

Since l *and* μ in expression (3.6.7) are the integers satisfying the condition, they are chosen such that the summation on the left hand side of (3.6.8) is approximately equal to α and $1-\alpha$ is roughly equally divided to the two tails. Therefore 100α-percent confidence interval for n may not be unique.

To determine a suitable value of N for computing a confidence interval for n, we adopt the scheme given below.

Let $s_j = \sum_{i=0}^{j} k_i$ (3.6.9)

be the jth partial sum of the sequence $k_0, k_1, k_2, \ldots$ for a given δ is defined to be the smallest integer such that $k_j \mid s_j < \delta$ (3.6.10)

Therefore, the required integer N is equal to $y_{max} + j - 1$ the criterion stated in the in equality (3.6.10) suggests that the posterior probabilities, beyond the value of N, will not contribute significantly.

Example: r=1

Suppose θ is known to be 0.2 the only success count shows that ten successes have been detected i.e. y=10. Hence, y_{max} is also equal to 10. When θ is known $\hat{n}$ = integer part of $y \mid \theta$ which is clearly sensible. Since y=10 and θ=0.2, $y \mid \theta$=10/0.2=50. Thus $\hat{n}$ =50.

Using the criterion (3.6.10) for $\delta = .005$, N is found to be 81. A 95-percent confidence interval for n is [30, 77] where the confidence coefficient, 95-percent is only approximation.

θ *is unknown:* When θ is unknown, assume that n and θ are independent. Let n have the same prior probability distribution as stated in (3.6.2). Suppose that the prior probability distribution for θ is in the form of beta distribution parameter v_1 and v_2. Let $P(\theta)$ denote the prior probability distribution of θ . Thus

$$P(\theta) \propto \theta^{v_1 - 1}(1-\theta)^{v_2 - 1}, 0 \le \theta \le 1 \tag{3.6.11}$$

which can represent a uniform prior $v_1 = v_2 = 1$ or conjugate prior representing information from a prior sample otherwise the joint posterior is then

$$P(n,\theta \mid y) \propto \theta^{t+v_1-1}(1-\theta)^{rn-t+v_2-1}\, h(n)\prod_{i=1}^{r}\frac{n!}{(n-x_i)!} \tag{3.6.12}$$

The marginal distribution of n can be obtained by integrating expression (3.6.12) with respect to θ from 0 to 1. Therefore,

$$P(n \mid y) \propto \frac{(rn-t+v_2-1)!}{(rn+v_1+v_2-1)!}\prod_{i=1}^{r}\frac{n!}{(n-x_i)!},\ for\ y_{\max} \le n \le N \tag{3.6.13}$$

Again the mode n of expression (3.6.13) would provide an estimate of n. similarly, if n= $y_{\max} + j$ for j= $0,1,2,\ldots N - y_{\max}$

$$P(n \mid y) = P(y_{\max} + j \mid y) = c\ k_j, \tag{3.6.14}$$

where $k_j = 1,\ \ if\ \ j = 0$

$$= k_{j-i}\prod_{i=0}^{r-1}\frac{[ry_{\max} - t + v_2 + (j-1)r + i]}{[ry_{\max} + v_1 + +v_2 + (j-1)r + i]}\frac{(y_{\max} + j)^r}{\prod_{i=1}^{r}(y_{\max} + j - y_i)},\ \ otherwise$$

is a recurrence formula for calculating k_j. The normalizing constant c in expression (3.6.14) can be computed in exactly the same manner as before. It should be noted that analytical results are easy to obtain using a suitable computer programs if necessary.

Clearly the distributions are such that consideration of small values of r is less difficult than consideration of large values.

Example: r=1 Suppose θ is unknown. An initial θ is found to be 0.2. Assume that the only success count gives ten successes. If δ in criterion (3.6.10) is chosen to be 0.005, then the estimates at various levels of certainty are presented in below table.

Estimates of n when r=1 and θ is unknown

	n	**N**	**95% confidence interval**
$v_1 = 2, v_2 = 5$	29 or 30	106	[17,102]
$v_1 = 5, v_2 = 17$	41 or 42	100	[23,96]
$v_1 = 10, v_2 = 37$	45 or 46	93	[26,89]
$v_1 = 20, v_2 = 77$	47 or 48	88	[27,83]

When the initial estimate of θ is made with high certainty, such as $v_1 = 20, v_2 = 77$ the point estimate of n is almost identical to the result given in example 1, in which θ is known. However, with θ unknown, confidence intervals are not as tight.

3.7 Bayes estimation of $1/\theta$:

Reliable estimation of $1/\theta$ is difficult when θ is close to 0, where a small change in θ will cause a large change in $1/\theta$. There is no unbiased estimator for$1/\theta$. This problem arises when estimating the size of certain animal population.

Suppose a lake contains an unknown number N of some species of fish. A random sample of size k is caught, tagged and released again. Later a sample of size n is obtained and the number Y of tagged fish in the sample is noted. Let us assume that each caught fish is immediately returned to the lake.

The n fish in the sample constitutes n Bernoulli trails with probability $\theta = \mathrm{k/N}$ of success. The population size N is k/θ.

Posterior distribution of θ is Beta ($\alpha + y, \beta - n - y$) the Bayes estimate under SELF is

$$E\left(\theta^{-1} \mid y\right) = \frac{B(\alpha + y - 1, \beta + n - y)}{B(\alpha + y, \beta + n - y)} = \frac{(\alpha + \beta + n - 1)}{(\alpha + y - 1)}$$

when prior is Beta (α, β).

If Haldanes nil prior is used $\alpha \to 0, \beta \to 0$;

$$E\left(\theta^{-1} \mid y\right) = (n-1)/(y-1), if \ \ y > 1 \text{and} \ \ \hat{N} = k(n-1)/(y-1), if \ \ y > 1$$

3.8 Lindely Approximation of Binomial Distribution:

We have $f(y|\theta)=\binom{n}{y}\theta^x(1-\theta)^{n-x}; x=0,1,2,...,n$ and $p(\theta)\propto\frac{1}{\theta(1-\theta)}$ be the prior distribution of θ

Also, $\mathrm{u}(\theta)=\theta$ then $u_1(\theta)=\frac{\partial u(\theta)}{\partial\theta}=1$

$$u_2(\theta)=\frac{\partial^2 u(\theta)}{\partial\theta^2}=0; \rho(\theta)\propto-[\log\theta]+\log(1-\theta)$$

Now, $L(\theta)=x\log\theta+(n-x)\log(1-\theta)+constant$

$$L_1(\theta)=\frac{x}{\theta}-\frac{(n-x)}{(1-\theta)} \text{ and } L_2(\theta)=\frac{-x}{\theta^2}-\frac{(n-x)}{(1-\theta)^2}, L_3(\theta)=\frac{2x}{\theta^3}-\frac{(n-x)}{(1-\theta)^3}$$

$$\rho_1(\theta)=\frac{\partial}{\partial\theta}[\rho(\theta)]$$

$$\Rightarrow \rho_1(\theta)=\frac{2\theta-1}{\theta(1-\theta)}$$

For an estimate of θ we have $L_1(\theta)=0$

$$\Rightarrow \frac{x}{\theta}-\frac{(n-x)}{(1-\theta)}=0,$$

$$\Rightarrow x(1-\theta)-(n-x)\theta=0,$$

$$\Rightarrow x-x\theta-n\theta+x\theta=0,$$

$$\Rightarrow \hat{\theta}=\frac{x}{n}$$

We have $L_2(\theta)=\frac{-x}{\theta^2}-\frac{(n-x)}{(1-\theta)^2}$

Then at $\theta=\hat{\theta}$

$$\Rightarrow L_2(\theta)|_{\theta=\hat{\theta}}=\frac{-n\hat{\theta}}{\hat{\theta}^2}-\frac{\left(n-n\hat{\theta}\right)}{\left(1-\hat{\theta}\right)^2}$$

$$\Rightarrow L_2(\theta)|_{\theta=\hat{\theta}}=\frac{-n\hat{\theta}\left(1-\hat{\theta}\right)^2-n\left(1-\hat{\theta}\right)\hat{\theta}^2}{\hat{\theta}^2\left(1-\hat{\theta}\right)^2}$$

$$\Rightarrow L_2(\theta)|_{\theta=\hat{\theta}}=\frac{-n\left(1-\hat{\theta}\right)-n\hat{\theta}}{\hat{\theta}\left(1-\hat{\theta}\right)}$$

$$\Rightarrow L_2(\theta)|_{\theta=\hat{\theta}} = \frac{-n}{\hat{\theta}(1-\hat{\theta})}$$

and $L_3(\theta)|_{\theta=\hat{\theta}} = \dfrac{-2n(1-2\hat{\theta})}{\hat{\theta}^2(1-\hat{\theta})^2}$

also, $\sigma^2 = \left[\dfrac{-\partial^2}{\partial\theta^2} L(\theta)\right]^{-1} = \left[\dfrac{x}{\theta^2} + \dfrac{n-x}{(1-\theta)^2}\right]^{-1}$

Then at $\theta = \hat{\theta}$

$$\Rightarrow \sigma^2 |_{\theta=\hat{\theta}} = \left[\frac{n\hat{\theta}}{\hat{\theta}^2} + \frac{n-n\hat{\theta}}{(1-\hat{\theta})^2}\right]^{-1}$$

$$\Rightarrow \sigma^2 |_{\theta=\hat{\theta}} = \left[\frac{\mathrm{n}\hat{\theta}(1-\hat{\theta})^2 + \mathrm{n}(1-\hat{\theta})\hat{\theta}^2}{\hat{\theta}^2(1-\hat{\theta})^2}\right]^{-1}$$

$$\Rightarrow \sigma^2 |_{\theta=\hat{\theta}} = \frac{\hat{\theta}(1-\hat{\theta})}{n}$$

Thus $E\left(\theta \mid \underset{\sim}{y}\right) = \dfrac{y}{n} + O\left(\dfrac{1}{n^2}\right)$

approximation is of $O\left(\dfrac{1}{n^2}\right)$

3.9 Normal approximation of binomial distribution:

Suppose $\underset{-}{Y} = y_1, y_2, ..., y_r$ be r independent random variables from binomial distribution with n known and θ is unknown parameter, then

$$f(Y \mid \theta) = \binom{n}{y} \theta^n (1-\theta)^{n-y}; 0 < \theta < 1; y = 0,1,2,\ldots,n$$

Then the likelihood function for $\mathrm{B}(\mathrm{n},\theta)$ is given by

$$L(\theta \mid y) = \prod_{i=1}^{r} \left[\binom{n}{y}\right] \theta^{\sum_{i=1}^{r} y_i} (1-\theta)^{nr - \sum_{i=1}^{r} y_i}$$

Let us consider the uniform prior i.e $P(\theta) \propto 1$

To construct the approximation, we need the second derivatives of the log-posterior density,

$$\log P(\theta \mid y) = \text{ constant} + \sum_{i=1}^{r} y_i \log\theta + \left(nr - \sum_{i=1}^{r} y_i\right) \log(1-\theta)$$

The first derivative is $\frac{\partial}{\partial\theta}\log P(\theta \mid y) = \frac{\sum_{i=1}^{r} y_i}{\theta} - \frac{(nr - \sum_{i=1}^{r} y_i)}{1-\theta}$

From which the posterior mode is readily obtained as $\hat{\theta} = \frac{\bar{y}}{n}$

The second derivative is $\frac{\partial^2}{\partial\theta^2}\log P(\theta \mid y) = -\left(\frac{r\bar{y}}{\theta^2} + \frac{(rn - r\bar{y})}{(1-\theta)^2}\right)$

and hence, the negative of hessian is $I(\theta) = \frac{-\partial^2 \log P(\theta \mid Y)}{\partial\theta^2} = \left[\frac{r\bar{y}}{\theta^2} + \frac{(nr - r\bar{y})}{(1-\theta)^2}\right]^{-1}$

and therefore we have

$$I(\hat{\theta}) = \frac{n^3 r}{\bar{y}(n-\bar{y})}$$

Therefore the large sample approximate posterior distribution is

$$P(\theta \mid Y) \approx N\left(\frac{\bar{y}}{n}, \frac{\bar{y}(n-\bar{y})}{n^3 r}\right)$$

Also, the 95% approximate HPD credible interval for θ under uniform prior i.e $P(\theta) \propto 1$ when n is known is

$$\left[\frac{\bar{y}}{n} - 1.96\sqrt{\frac{\bar{y}(n-\bar{y})}{n^3 r}}, \frac{\bar{y}}{n} + 1.96\sqrt{\frac{\bar{y}(n-\bar{y})}{n^3 r}}\right]$$

Example 3.1(Hoff, P.D. (2009): Each female of age 65 or over in1998 General social survey was asked whether or not they were generally happy. Let yi=1,if the respondent i reported being generally happy & let yi=0 otherwise.Suppose 129 individuals were surveyed out of which 118 individuals reported being generally happy & remaining 11 do not report being generally happy. We have developed various programs for MLE and Bayes estimates of this example in R software and the results are given in table 3.1.1.

 # Mle of binomial distribution is given by

```
library(stats4)
binomNLL <- function(theta, y, n) {
 -sum(dbinom(y, prob = theta, size = n, log = TRUE))
 }
```

```
y<-118
est <- optim(par = 0.5, fn = binomNLL, n=129, method = "BFGS",y=y)
est
$par
[1] 0.9147252
$value
[1] 2.080956
$counts
function gradient
     51       9
$convergence
[1] 0
```

 # Posterior mean, variance and credible interval of binomial distribution under natural conjugate prior

```
Post.mav<-function(a,b,y,n){
 pm.theta<-(a+y)/(a+b+n)
 pvar.theta<-(a+y)*(b+n-y)/((a+b+n)*(a+b+n+1)^2)
 ci<-qbeta(c(0.025,0.975),a+y,b+n−y)
 list(Posterior.mean=pm.theta,Posterior.variance=pvar.theta,Credible.interval=ci)
 }
> Post.mav(1,1,118,129)
$Posterior.mean
[1] 0.9083969
$Posterior.variance
[1] 0.0006256177
$Credible.interval
[1] 0.8536434 0.9513891
```

 # Posterior mean,variance and credible interval of binomial distribution under Jeffrey’s prior

```
Post.mav<-function(y,n)
 {
 pm.theta<-(y+0.5)/(n+1)
```

```
pvar.theta<-(y+0.5)*(n-y+0.5)/((n+2)*(n+1)^2)
ci<-qbeta(c(0.025,0.975),y+0.5,n−y+0.5)
list(Posterior.mean=pm.theta,Posterior.variance=pvar.theta,Credible.interval=ci)
}
> Post.mav(118,129)
$Posterior.mean
[1] 0.9115385
$Posterior.variance
[1] 0.0006155427
$Credible.interval
[1] 0.8572894 0.9538477
```

 # Posterior mean, variance and credible interval of binomial distribution under ALIP prior

```
Post.mav<-function(y,n)
{
pm.theta<-y/n
pvar.theta<-y*(n-y)/((n+1)*n^2)
ci<-qbeta(c(0.025,0.975),y,n−y)
list(Posterior.mean=pm.theta,Posterior.variance=pvar.theta,Credible.interval=ci)
 }
> Post.mav(118,129)
$Posterior.mean
[1] 0.9147287
$Posterior.variance
[1] 0.0006000009
$Credible.interval
[1] 0.8610175 0.9563177
```

Table 3.1.1: Posterior mean, variance and credible interval of binomial distribution under different priors.

Type of Prior	Posterior Mean	Posterior variance	Credible Interval
Conjugate Prior	0.9083969	0.0006256177	0.8536434,0.9513891
Jeffrey's Prior	0.9115385	0.0006155427	0.8572894,0.9538477
ALIP	0.9147287	0.0006000009	0.8610175,0.9563177

Graphical representation of Beta Posterior under two different sample sizes in S-Plus:

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
par(mfrow=c(2,2))
theta<-seq(0,1,length=100)
a<-1; b<-1
n<-5 ; y<-1
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p("),theta,"|y)",sep="")),
xlab=expression(theta),lwd=2)
mtext(expression(paste("beta(1,1) prior, ",
italic("n"),"=5",italic(sum(y[i])),"=1",sep="")))
 #abline(v=c((a+y-1)/(a+b+n-
2),(a+y)/(a+b+n)),col=c("black","gray"),lty=c(2,2))
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=2)
legend(.45,2.4,legend=c("prior","posterior"),lwd=c(2,2),c
ol=c("gray","black"), bty="n")
a<-3; b<-2
n<-5 ; y<-1
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p("),theta,"|y)",sep="")),
xlab=expression(theta),lwd=2)
# expression(italic(paste("p(",theta,"|y)",sep=""))),
xlab=expression(theta),lwd=2)
```

```
mtext(expression(paste("beta(3,2) prior, ",
italic("n"),"=5 ",italic(sum(y[i])),"=1",sep="")))
#abline(v=c((a+y-1)/(a+b+n-2), (a+y)/(a+b+n)) ,
col=c("green","red"))
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=2)
a<-1 ; b<-1
n<-100; y<-20
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p("),theta,"|y)",sep="")),
xlab=expression(theta),lwd=2)
# expression(italic(paste("p(",theta,"|y)",sep=""))),
xlab=expression(theta),lwd=2)
mtext(expression(paste("beta(1,1) prior, ",
italic("n"),"=100 ",italic(sum(y[i])),"=20",sep="")))
 #abline(v=c((a+y-1)/(a+b+n-2), (a+y)/(a+b+n)) ,
col=c("green","red") )
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=2)
a<-3 ; b<-2
n<-100; y<-20
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p("),theta,"|y)",sep="")),
xlab=expression(theta),lwd=2)
#
expression(italic(paste("p(",theta,"|y)",sep=""))),xlab=e
xpression(theta),lwd=2)
mtext(expression(paste("beta(3,2) prior, ",
italic("n"),"=100 ",italic(sum(y[i])),"=20",sep="")))
#abline(v=c((a+y-1)/(a+b+n-2), (a+y)/(a+b+n)) ,
col=c("green","red"))
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=2)
dev.off()
```

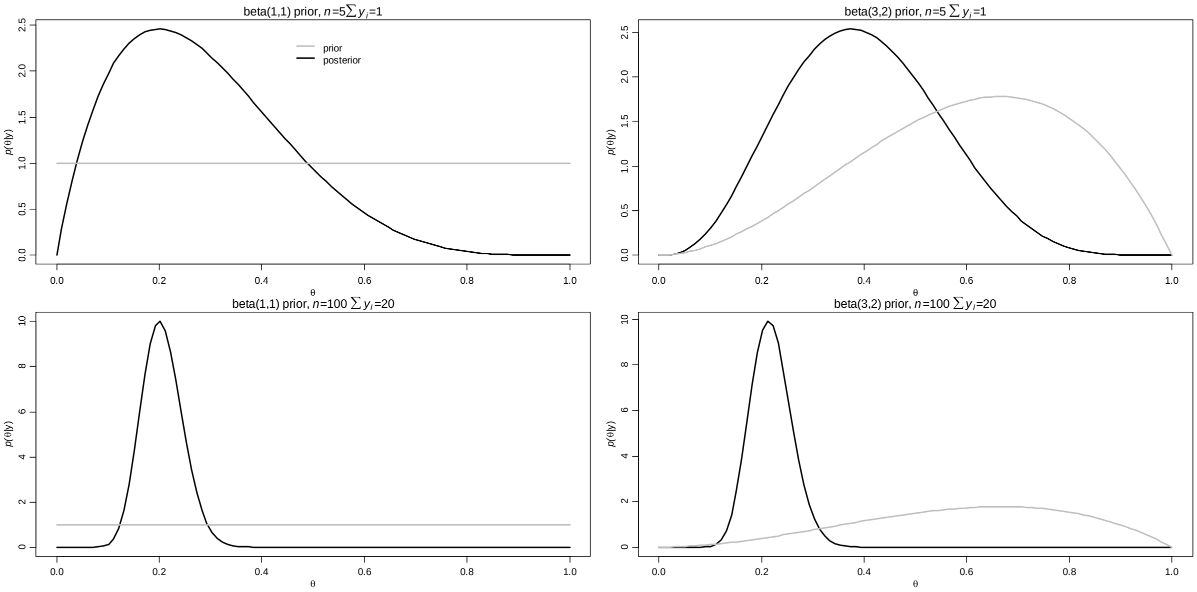

Figure: 3.1

 # Graphical representation of Posterior distribution and 95% credible interval under different Priors.

```
a<-1 ; b<-1 #prior
n<-129 ; y<-118 #data
theta.support<-seq(0.8,1.0,length=1500)
plot(theta.support, dbeta(theta.support, a+y, b+n-y),
type="l",xlab="theta",
ylab="p(theta|y)" )
qbeta( c(.025,.975), a+y,b+n-y)
abline(v=qbeta( c(.025,.975), a+y,b+n-y))
```

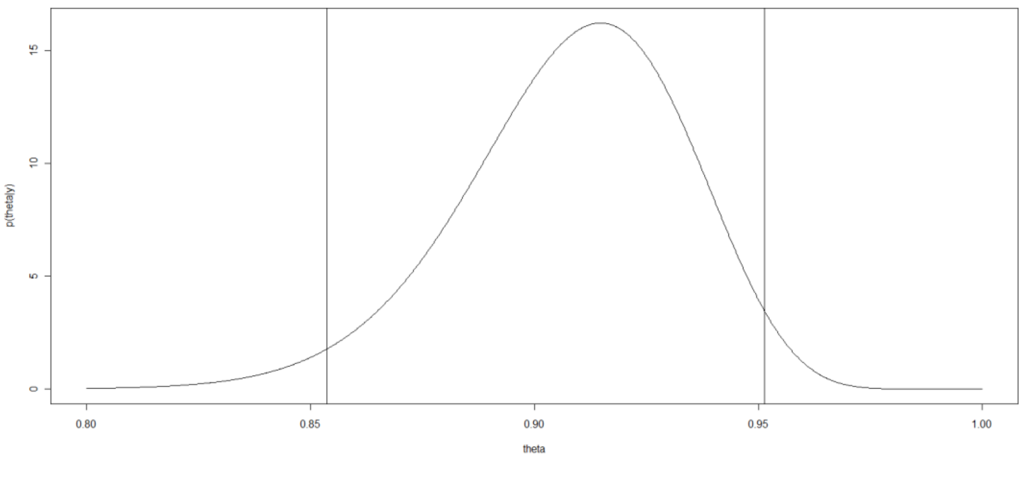

Figure: 3.2

```
a<-1 ; b<-1 #prior
n<-129 ; y<-118 #data
theta.support<-seq(0.8,1.0,length=1500)
plot(theta.support, dbeta(theta.support, y+0.5, n-y+0.5),
type="l",xlab="theta",
ylab="p(theta|y)" )
qbeta(c(0.025,0.975),y+0.5,n-y+0.5)
abline(v=qbeta( c(.025,.975), a+y,b+n-y))
```

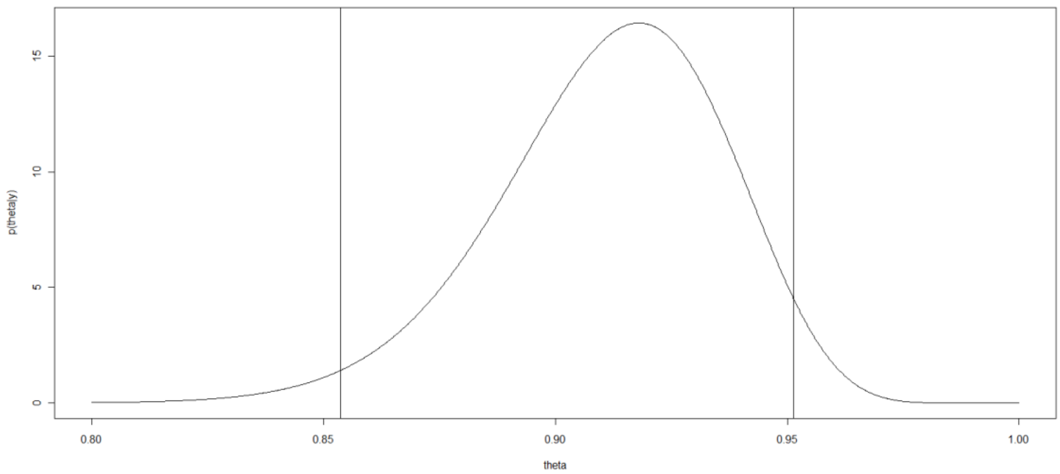

Figure 3.3

```
a<-1 ; b<-1 #prior
n<-129 ; y<-118 #data
theta.support<-seq(0.8,1.0,length=1500)
plot(theta.support, dbeta(theta.support, y+0.5, n-y+0.5),
type="l",xlab="theta",
ylab="p(theta|y)" )
qbeta(c(0.025,0.975),y,n-y)
abline(v=qbeta( c(.025,.975), a+y,b+n-y))
qbeta(c(0.025,0.975),y,n-y)
```

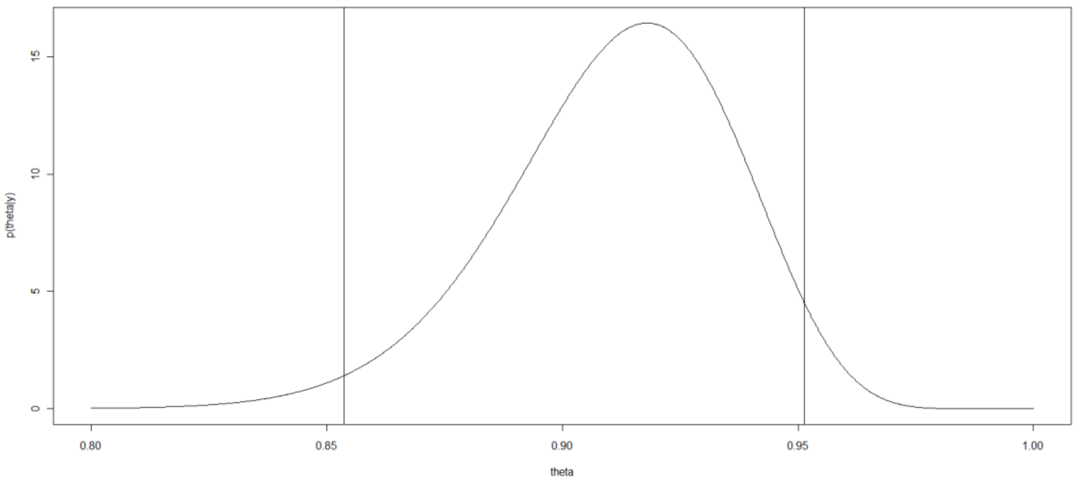

Figure 3.4

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
a<-1;b<-1#Prior
n<-129;y<-118
theta.support<-seq(0.8,1.0,length=5000)
plot(theta.support, dbeta(theta.support, a+y, b+n-y),
type="l", xlab=expression(theta),
      ylab=expression(paste(italic("p("),theta,"|y)")))
pth<-dbeta(theta.support, a+y, b+n-y)
pth<-pth
ord<- order(-pth)
xpx<-cbind(theta.support[ord], pth[ord])
xpx<-cbind(xpx,cumsum(xpx[,2])/sum(xpx[,2]))
hpd<-function(x,dx,p){
md<-x[dx==max(dx)]
px<-dx/sum(dx)
pxs<--sort(-px)
ct<-min(pxs[cumsum(pxs)< p])
list(hpdr=range(x[px>=ct]),mode=md) }
tmp<-hpd(xpx[,1],xpx[,2],.5)$hpdr
lines( x=c(tmp[1],tmp[1],tmp[2],tmp[2]),
y=dbeta(c(0,tmp[1],tmp[2],0),a+y,b+n-y)
,col=gray(.75),lwd=2 )
tmp<-hpd(xpx[,1],xpx[,2],.75)$hpdr
```

```
lines( x=c(tmp[1],tmp[1],tmp[2],tmp[2]),
       y=dbeta(c(0,tmp[1],tmp[2],0),a+y,b+n-y)
,col=gray(.5),lwd=2 )
tmp<-hpd(xpx[,1],xpx[,2],.95)$hpdr
lines( x=c(tmp[1],tmp[1],tmp[2],tmp[2]),
      y=dbeta(c(0,tmp[1],tmp[2],0),a+y,b+n-y)
,col=gray(0),lwd=2 )
tmp<-qbeta( c(.025,.975), a+y,b+n-y)
lines( x=c(tmp[1],tmp[1],tmp[2],tmp[2]),
y=dbeta(c(0,tmp[1],tmp[2],0),a+y,b+n-y)
,col=gray(0),lwd=2 ,lty=2 )
legend(0.95, 14, c("50% HPD","75% HPD","95% HPD","95%
quantile-based"),
col=c(gray(.75),gray(.5),gray(0),gray(0)),lty=c(1,1,1,2),
lwd=c(2,2,2,2),bty="n")
```

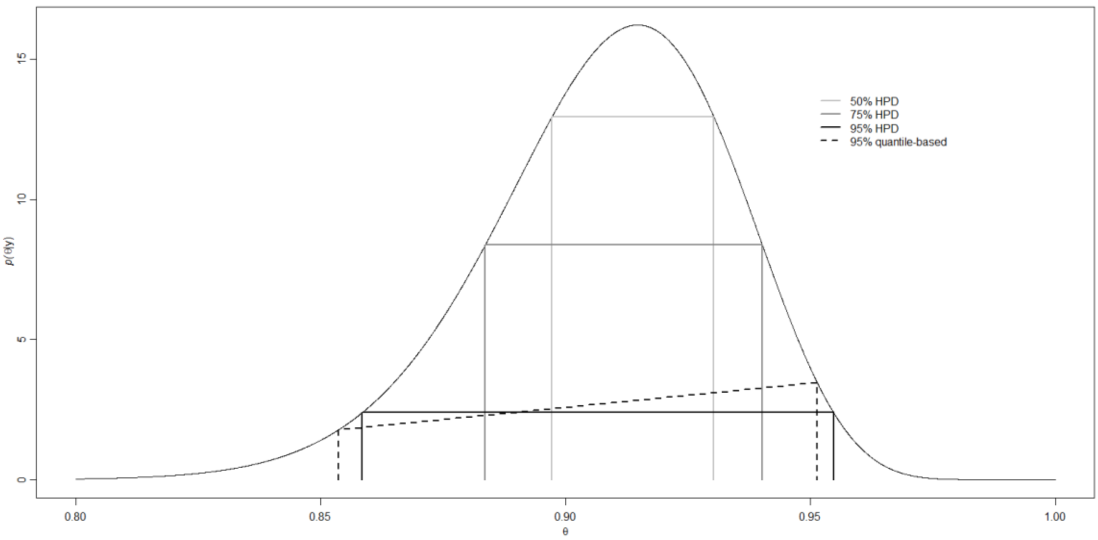

Figure 3.5

Graphical representation of binomial distribution with different values of n and θ in S-PLUS

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,2))
n<-10
theta<-.2
```

```
 plot(0:n,dbinom(0:n,n,theta),
type="h",lwd=2,xlab=expression(italic(y)),

ylab=expression(paste("Pr(",italic("Y=y"),"|",theta==.2,i
talic(", n="),"10)",sep="")))
 #MTEXT(EXpression(
 # italic(paste("n=",10,", ",theta==0.2))),side=3,cex=.8)
 n<-10
 theta<-.8
 plot(0:n,dbinom(0:n,n,theta),
type="h",lwd=2,xlab=expression(italic(y)),

ylab=expression(paste("Pr(",italic("Y=y"),"|",theta==.8,i
talic(", n="),"10)",sep="")))
 #mtext(expression(
 # italic(paste("n=",10,", ",theta==0.8))),side=3,cex=.8)
```

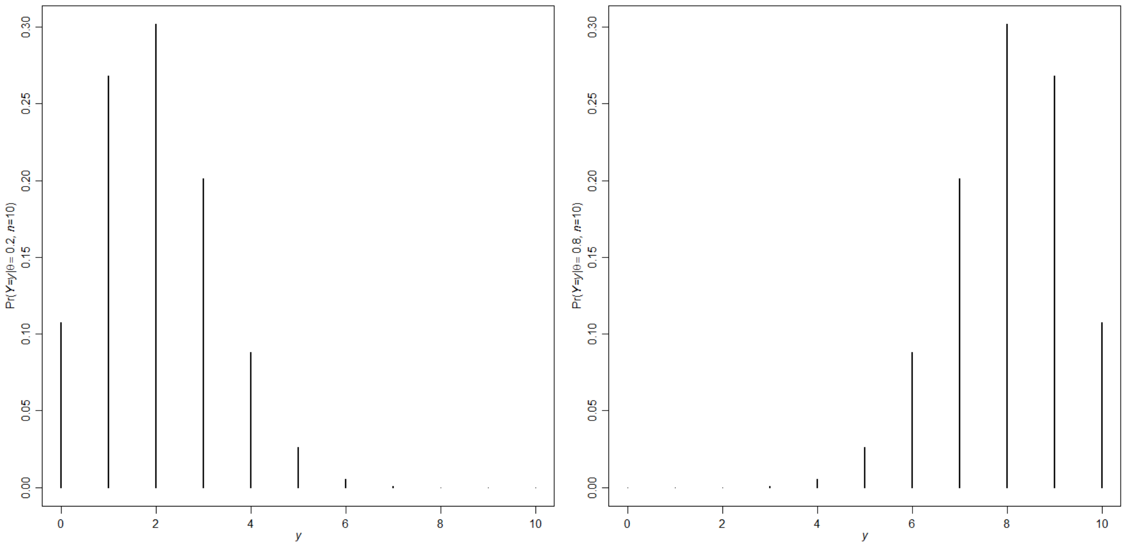

Figure 3.6

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
 par(mfrow=c(1,2))
 n<-100
 theta<-.2
 plot(0:n,dbinom(0:n,n,theta), type="h",lwd=2,xlab=expression(italic(y)),
 ylab=expression(paste("Pr(",italic("Y=y"),"|",theta==.2,italic(", n="),"100)",sep="")))
```

```
n<-100
theta<-.8
plot(0:n,dbinom(0:n,n,theta), type="h",lwd=2,xlab=expression(italic(y)),
ylab=expression(paste("Pr(",italic("Y=y"),"|",theta==.8,italic(", n="),"100)",sep="")))
```

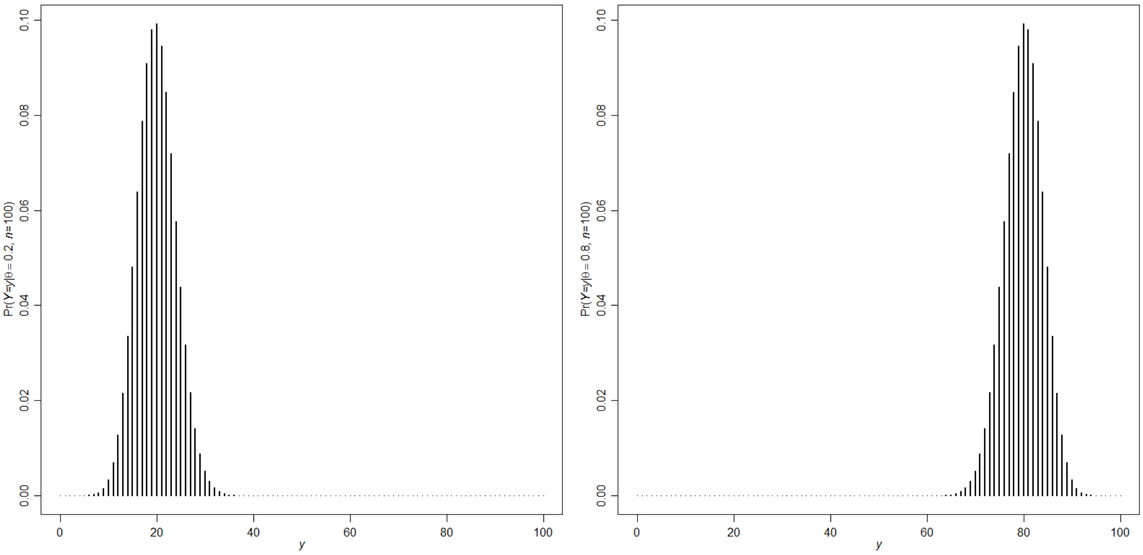

Figure 3.7

CHAPTER – 4

BAYESIAN ESTIMATION FOR POISSON DISTRIBUTION

4.1 Introduction:

Poisson distribution was discovered by the French mathematician and Physicist Simeon Denis Poisson (1781-1840) who published it in 1837. Poisson distribution is a limiting case of the binomial distribution under the following conditions:

i) n, the number of the trials is indefinitely large, i.e. $n \to \infty$

ii) The constant probability of success for each trial is indefinitely, i.e. $\theta \to 0$

iii) $n\theta = \lambda$ (say) is finite

Thus $\theta = \frac{\lambda}{n}$, $1-\theta = 1-\frac{\lambda}{n}$ where λ is a positive real number. The probability of y in a series of n independent trial is:

$$f(y;n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y} \quad , y = 0,1,2,\ldots,n \qquad (4.1.1)$$

We want the limiting form of (4.1.1) under above conditions. Hence

$$f(y;n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

$$f(y;n,\theta) = \binom{n}{y}\left(\frac{\theta}{1-\theta}\right)^{y}(1-\theta)^{n}$$

$$= \frac{n(n-1)(n-2)\ldots\ldots(n-x+1)}{y!}\frac{(\lambda/n)^{y}}{(1-\lambda/n)^{y}}(1-\lambda/n)^{n}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\ldots\ldots\left(1-\frac{x-1}{n}\right)}{y!\left(1-\frac{\lambda}{n}\right)^{y}}\lambda^{y}\left(1-\frac{\lambda}{n}\right)^{n}$$

$$\therefore \lim_{n\to\infty} f(y;n,\theta) = \frac{e^{-\lambda}\lambda^{y}}{y!}; \quad y = 0,1,2,\ldots\ldots$$

which is the required probability function of Poisson distribution. λ is known as parameter of Poisson distribution. Thus a random variable Y is said to follow a Poisson distribution of it assumes only non-negative values and its probability mass function is given by

$$f(y\,|\,\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}; \quad y = 0,1,2,\ldots\ldots;\lambda > 0$$

4.2 Maximum likelihood estimate for Poisson distribution:

The likelihood function of the Poisson distribution is given as

$$L(\lambda \mid Y) = \frac{e^{-n\lambda} \ \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y!}$$

Applying log on both sides of above equation we have

$$\log L(\lambda \mid Y) = -n\lambda + \sum_{i=1}^{n} y_i \log\lambda + constant$$

The maximum likelihood estimate of λ (called $\hat{\lambda}$ hereafter) is obtained by taking the derivative of

$$\log L(\lambda \mid Y) = -n\lambda + \sum_{i=1}^{n} y_i \log\lambda + \text{constant}$$

Differentiating with respect to λ and finally setting the derivative equal to zero and solving for λ.

i.e. $$\frac{\partial \log L(\lambda \mid Y)}{\partial\lambda} = 0$$

$$\Rightarrow \ -n + \frac{1}{\lambda}\sum_{i=1}^{n} y_i = 0$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

In order to ascertain this is indeed the maximum likelihood estimate, we would also take the second order derivative. It can be shown that the second order derivative satisfies the criteria for global optimality for λ.

4.3 Bayes Estimator for Poisson distribution:

The pmf of Poisson distribution is

$$f(Y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0,1,2,........ : \lambda > 0$$

Where λ is now treated as a random variable. A straight forward computation gives Fisher information. $I(\lambda) = \frac{1}{\lambda}$

Hence Jeffrey's prior $P(\lambda) \propto \lambda^{\frac{-1}{2}}$ which an improper (or quasi) prior since

$$\int_0^{\infty} P(\lambda)d\lambda \neq 1$$

Let us consider a more general class of priors,

$$P(\theta) \propto \left(\frac{1}{\lambda^{\frac{c}{2}}}\right), \quad c \geq 0 \text{ we have}$$

$$L(\lambda \mid Y) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}, \quad where\ Y = (y_1, y_2, \ldots, y_n)$$

The posterior distribution of λ is given by

$$P(\lambda \mid Y) = \frac{k\ e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\lambda^{\frac{c}{2}}},$$

Where $k^{-1} = \int_0^{\infty} e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i - \frac{c}{2}}\, d\lambda$

$$k^{-1} = \frac{\Gamma\left(\sum_{i=1}^{n} y_i - \frac{c}{2} + 1\right)}{(n)^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}}$$

Thus $P(\lambda \mid Y) = \dfrac{(n)^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}}{\Gamma\left(\sum_{i=1}^{n} y_i - \frac{c}{2} + 1\right)} e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i - \frac{c}{2}}$,

and the Bayes estimator of λ is

$$\hat{\lambda} = E(\lambda \mid Y) = \int_0^{\infty} \frac{(n)^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}}{\Gamma\left(\sum_{i=1}^{n} y_i - \frac{c}{2} + 1\right)} e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}\, d\lambda$$

$$\hat{\lambda} = \frac{(n)^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}}{\Gamma\left(\sum_{i=1}^{n} y_i - \frac{c}{2} + 1\right)} \frac{\Gamma\left(\sum_{i=1}^{n} y_i - \frac{c}{2} + 2\right)}{(n)^{\sum_{i=1}^{n} y_i - \frac{c}{2} + 2}}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i - \frac{c}{2} + 1}{n}$$

For c=1, $P(\lambda) \propto \lambda^{\frac{-1}{2}}$ (Jeffery's prior) we have

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{n} y_i + \frac{1}{2}}{n}$$

For c=2, $P(\lambda) \propto \frac{1}{\lambda}$, we have

$$\hat{\lambda}_2 = \bar{y}$$

which is the maximum likelihood as well as the unique minimum variance unbiased estimator (UMVUE) of λ.

For c=0, $P(\lambda) \propto 1$ (uniform prior), we have

$$\hat{\lambda}_3 = \frac{\sum_{i=1}^{n} y_i + 1}{n}$$

We note that for n quite large as compared to $c, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ well all be numerically very close to each other. Thus, the effect of the prior distribution on actual estimators is rather small when the sample size is large this fact is well exhibited by equation

$$P(\theta \mid Y) \propto P(\theta) L(y_1, y_{2,}, \ldots\ldots, y_n \mid \theta)$$

which shows that at least in large samples the posterior is more dominated by likelihood $\prod_{i=1}^{n} f(y_i \mid \theta) = L(y_1, y_2, \ldots, y_n \mid \theta)$ than the prior $P(\theta)$. Moreover, we can show that $E\left[\left(\hat{\lambda}_1 - \hat{\lambda}_2\right)^2\right] \to 0\, as\ n \to 0$ for any fixed c_1 and c_2. By chebychev's inequality it follows that for any $\varepsilon > 0$

$$\lim_{n \to \infty} P\left\{\left|\lambda_{c_1} - \lambda_{c_2}\right| < \varepsilon\right\} = 1.$$

Showing thereby that in large samples the choice of the constant c (i.e. the choice of the prior distribution is not very crucial).

4.4 Comparison of non-informative priors for number of defects (possion) model:

Now, we consider the Bayesian analysis of the model for the number of defects. We considered the two non-informative priors (Jeffrey and uniform) and will study their performance using different distribution performance measures. The posterior distribution and posterior productive distribution for the parameter of the model for the number of defects will also be derived using the above said prior.

Posterior distribution of parameter using Jeffery prior (JP) usually, the distribution of the discrete time-to failure system follows the Poisson distribution so the probability mass function (pmf) of the Poisson distribution for a random variable Y having parameter λ is

$$f(Y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}; \quad y = 0,1,2,.... \ ; \lambda > 0 \tag{4.4.1}$$

The likelihood function for a simple random of size n is given by

$$i.e. \quad L(\lambda \mid Y) = \prod_{i=1}^{n} f(y_i) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} \tag{4.4.2}$$

Here the parameter λ is unknown. In the situations where one does not have much information about the parameter, Jeffrey (1946, 1961) suggested a non- informative prior (NIP). This defines the density of the parameter proportional to the square root of the determinant of the Fisher information matrix. Symbolically the Jeffrey prior distribution $p_j(\lambda)$ is given by $p_j(\lambda) \propto \sqrt{I(\lambda)}$ where $I(\lambda) = -E\left[\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2}\right]$ when there are multiple parameters, I is the fisher information matrix, and is given as $I_{ij}(\lambda) = -E\left[\frac{\partial^2 \log L(\lambda)}{\partial \lambda_i \, \partial \lambda_j}\right]$.

In this case the Jeffrey prior becomes

$$p_j(\lambda) \propto \sqrt{\det[I(\lambda)]} \tag{4.4.3}$$

The likelihood function from (4.4.2) is

$$L(\lambda \mid Y) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

$$\log L(\lambda \mid y) = -n\lambda + \sum_{i=1}^{n} Y_i \log \lambda + cons\tan t$$

Taking the derivative with respect to λ

$$\frac{\partial \log L(\lambda \mid Y)}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i \qquad (4.4.4)$$

Setting the derivate equal to 0 and finally solving for λ we get

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$$

Which is the likelihood estimate of λ.

Then $I_{ij}(\lambda)$ is given by

$$I_{ij}(\lambda) = -E\left[\frac{\partial^2 \log L(y, \lambda)}{\partial \lambda^2}\right] = \frac{1}{\lambda} \qquad (4.4.5)$$

The Jeffrey prior for a Poisson distribution with parameter λ is $p_j(\lambda) \propto \lambda^{-1/2}$.

The posterior distribution of parameter λ for given data $(Y = y_1, y_2, ..., y_n)$ using equation (4.4.2) and (4.4.5) is

$$p(\lambda \mid Y) \propto L(y, \lambda) P_i(\lambda)$$

$$\Rightarrow\ p(\lambda \mid Y) \propto \lambda^{\sum_{i=1}^{n} y_i - 1/2} e^{-n\lambda}$$

$$\Rightarrow p(\lambda \mid Y) = k\, e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i - 1/2} \qquad (4.4.6)$$

Where k is normalizing constant and given by

$$k^{-1} = \int_0^{\infty} e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i + 1/2 - 1} d\lambda$$

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

Then from equation (4.4.6)

$$p(\lambda \mid Y) = \frac{n^{\sum_{i=1}^{n} y_i + \frac{1}{2}} e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i - 1/2}}{\Gamma\left(\sum_{i=1}^{n} y_i + \frac{1}{2}\right)} \tag{4.4.7}$$

which is the density function of gamma distribution with parameter n and $\sum_{i=1}^{n} y_i + \frac{1}{2}$.

So the posterior distribution of λ given data is gamma (n, $\sum_{i=1}^{n} y_i + \frac{1}{2}$).

The following data of size 10 is generated from Poisson distribution with parameter $\lambda = 2!$ 1, 2, 3,1,2,0,5,3,1 and 2. The sum of all the 10 observations is 20 ($i.e. \sum_{i=1}^{n} y_i = 20, n = 10$). So the posterior distribution of parameter λ for the given data $(Y = y_1, y_2, ..., y_{10})$, using equation (4.4.6) is the gamma distribution with parameters $\alpha = 10 and \beta = 20.50$ and i.e. gamma (10, 20.50).

4.5 The posterior distribution of the parameter using the uniform prior (UP):

Laplace (1774, 1812) found that it worked exceptionally well to simply choose always the prior for λ to be constant $[P(\lambda) = 1]$ on the parameter space. The uniform prior (non-informative prior) distribution of

$$P(\lambda) \propto 1, \quad 0 < \lambda < \infty \tag{4.5.1}$$

The posterior distribution of parameter λ for given data $(\mathrm{Y} = \mathrm{y}_1, \mathrm{y}_{2,} ......, \mathrm{y}_\mathrm{n})$ using equation (4.4.2) and (4.4.3) is

$$P(\lambda \mid Y) \propto \lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}$$

$$P(\lambda \mid Y) = k \lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda} \tag{4.5.2}$$

Where $k^{-1} = \int_0^{\infty} \lambda^{\sum_{i=1}^{n} y_i + 1 - 1} e^{-n\lambda} d\lambda$

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

From equations (4.5.2) we have

$$P(\lambda \mid Y) = \frac{n^{\sum_{i=1}^{n} y_i + 1}}{\Gamma\left(\sum_{i=1}^{n} y_i + 1\right)} \lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda} \quad (4.5.3)$$

which is the density function of Gamma distribution with parameters $n, \sum_{i=1}^{n} y_i + 1$. The posterior mode $\frac{\sum_{i=1}^{n} y_i}{n}$, of the gamma distribution using uniform prior is equal to its classical counter parts the maximum likelihood estimator (MLE) and the uniformly minimum variance unbiased estimator (UMVUE). Hence for the data considered above, the posterior distribution of the parameter λ for given data $(Y = y_1, y_2, ..., y_{10})$ using equations (4.5.3) is a gamma distribution with parameters $\alpha = 10.00, \beta = 21.00$ i.e. gamma (10, 21).

4.6 Comparison of non informative priors with respect to posterior variance:

The posterior variances of parameter λ using two non informative priors are given in the following table

Posterior Variance of Parameter λ

	Variance using	
	NIP	
	JP	UP
λ	0.2050	0.2100

From the above table it is obvious that var(λ) using Jeffrey prior is equal to var(λ) using uniform prior. That is both priors are approximately equally efficient. So either of them can be used as a non informative prior and hence robustness with respect to the choice of non-informative prior is observed. However it is reasonable to prefer the uniform prior being simpler as compared to Jeffrey prior.

4.7 Comparison based on Bayesian point estimates:

The Bayesian estimates of λ are presented in table classical counterparts are also given in the table. From table we conclude that both posterior mode and posterior mode posterior mean using the two priors are almost same as the MLE and UMVUE.

Bayesian point estimates using NI priors:

Bayesian estimates		Classical counterpart (MLE UMVUE)
JP	2.05	2.00
UP	2.10	2.00
JP	1.95	2.00
UP	2.00	2.00

4.8 Comparison of prior using coefficient of Skewness:

This section provides the comparison of priors using coefficient of Skeweness. The coefficient of Skewness is calculated from the posterior distributions and is discussed below. The coefficient of the posterior distribution is given by $y_1 = 2\sqrt{\frac{1}{\beta}}$.

Coefficient of Skewness for posterior distribution:

	Posterior parameters	**Coefficient of skewness**
	(α,β)	γ_1
JP	(10.00,20.50)	0.4417
UP	(10.00,21.50)	0.4365

From above table we observe that $\gamma_1 > 0$, therefore the posterior distribution based on the Jeffrey's and the uniform priors are not symmetrically, rather they both are slightly positively and almost equally skewed. However, because of the simplicity the uniform prior may be preferred to the Jeffrey's prior.

4.9 Comparison of priors using Bayes estimator:

Bayes decision is a decision $'d^*'$ which minimizes risk function and $'d^*'$ is the best decision. If decision is the choice of the estimator then the Bayes decision is the Bayes estimator. The below given Bayes estimator based on non informative prior for different loss function.

Loss function $L(\lambda, Y)$	Bayes estimator d^*	Prior	Posterior parameters (α, β)	Bayes estimator d^*	Classical estimates
$L_1 = \left(1 - \frac{d}{\lambda}\right)^2$	$\frac{\beta - 2}{\alpha}$	Jeffery's prior	(10,20.50)	1.85	2.00
		Uniform prior	(10,21)	1.90	2.00
$L_2 = \left(\frac{\lambda - d}{\lambda}\right)^2$	$\frac{\beta - 1}{\alpha}$	JP	(10,20.50)	1.95	2.00
		UP	(10,21)	2.00	2.00
$L_3 = (\lambda - d)^2$	$\frac{\alpha}{\beta}$	JP	(10,20.50)	2.05	2.00
		UP	(10,21)	2.10	2.00

From the above table we see that Bayes estimator for different loss function $L_1, L_2, \text{and} L_3$ using two priors are almost equal to UMVUE and MLE.

4.10 On the double prior selection for the parameter of Poisson distribution:

Sometimes it may happen that for a single true unknown parameter, different prior information is available; usually we use one informative prior to incorporate that prior knowledge and ignoring the other information. So to include two different kind of information in the analysis, two different priors are selected for a single unknown parameter of Poisson distribution. Here we will make use of three double priors namely Gamma-chi-square distribution, gamma-exponential distribution, chi-square-exponential distribution and one as prior: Gamma distribution for the unknown parameter of the Poisson model.

Let $(Y = y_1, y_{2,} ......, y_n)$ be a random-sample, drawn from the passion distribution having unknown parameter λ. The likelihood function of the sample observations $(Y = y_1, y_{2,} ......, y_n)$ is

$$L(\lambda \mid Y) = \prod_{i=1}^{n} f(y_i, \lambda) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}, \; y = 0,1,2,.....$$

where $\lambda > 0$ is unknown parameter.

4.11 Posterior distribution of the unknown parameter of Poisson distribution under Gamma-chi-square distribution as a double prior:

It is assumed that the prior distribution of λ is Gamma distribution with hyper parameters $'a_1' and 'b_1'$ which is given below;

$$P_{11}(\lambda) = \frac{(b_1)^{a_1} e^{-\lambda b_1} \lambda^{a_1 - 1}}{\Gamma a_1}, \lambda > 0, a_1, b_1 > 0 \tag{4.11.1}$$

Similarly, the second prior distribution is assumed to be the chi-square distribution with hyper parameter c_1. The pdf of the prior is;

$$P_{12}(\lambda) = \frac{(1/2)^{\frac{c_1}{2}} e^{-\lambda/2} \lambda^{\frac{c_1}{2}-1}}{\Gamma \frac{c_1}{2}}, \lambda > 0, c_1 > 0 \qquad (4.11.2)$$

Now we define the double prior for λ by combining these two priors which is as follows;

$$P_1(\lambda) \propto P_{11}(\lambda) P_{12}(\lambda)$$

$$\Rightarrow P_1(\lambda) = k\left(\lambda^{a_1-1} e^{-\lambda b_1} e^{\frac{-\lambda}{2}} \lambda^{\frac{c_1}{2}-1}\right)$$

Where $k = \dfrac{b_1^{\,a_1}}{\Gamma a_1 \Gamma c_1 2^{\frac{c_1}{2}}} = \dfrac{b_1^{\,a_1}(0.5)^{0.5c_1}}{\Gamma a_1 \Gamma(0.5 c_1)}$

Now the posterior distribution of λ for given data 'Y' is

$$p_1(\lambda \mid Y) \propto L(\lambda \mid Y) P_1(\lambda)$$

$$\Rightarrow p_1(\lambda \mid Y) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i} \lambda^{a_1 + \frac{c_1}{2} - 1 - 1} e^{-\lambda\left(b_1 + \frac{1}{2}\right)}$$

$$\Rightarrow p_1(\lambda \mid Y) \propto \lambda^{(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1) - 1} e^{-\lambda\left(n + b_1 + \frac{1}{2}\right)}, \lambda > 0$$

$$\Rightarrow p_1(\lambda \mid Y) = k\lambda^{(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1) - 1} e^{-\lambda\left(n + b_1 + \frac{1}{2}\right)}, \lambda > 0 \qquad (4.11.3)$$

$$where \;\; k^{-1} = \int_0^{\infty} \lambda^{(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1) - 1} e^{-\lambda\left(n + b_1 + \frac{1}{2}\right)} d\lambda$$

Hence from equation (4.11.3) we have

$$p_1(\lambda \mid Y) = \frac{\left(n + b_1 + \frac{1}{2}\right)^{a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1}}{\Gamma\left(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1\right)} e^{-\lambda\left(n + b_1 + \frac{1}{2}\right)} \lambda^{(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1) - 1} \qquad (4.11.4)$$

Hence the posterior distn of λ for data is Gamma distribution with parameter

$$\alpha_1 = \left(n + b_1 + \frac{1}{2}\right), \; and \; \beta_1 = \left(a_1 + \frac{c_1}{2} + \sum_{i=1}^{n} y_i - 1\right)$$

4.12 Gamma exponential distribution as a double prior:

It is assumed that the double prior distribution of λ is Gamma distribution with hyper parameter $'a_2'$and$'b_2'$and exponential distribution with hyper parameter c_2 which is given below:

$$P_2(\lambda) \propto \lambda^{a_2-1} e^{-\lambda b_2} e^{-\lambda c_2}, \lambda > 0 \qquad (4.12.1)$$

Now the posterior distribution of λ for the given data 'Y' is:

$$p_2(\lambda \mid Y) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i} e^{-\lambda b_2} \lambda^{a_2-1} e^{-\lambda c_2}$$

$$\Rightarrow p_2(\lambda \mid Y) = k\, e^{-\lambda(b_2+n+c_2)} \lambda^{(a_2+\sum_{i=1}^{n} y_i)-1}, \lambda > 0 \qquad (4.12.2)$$

$$where \;\; k^{-1} = \int_0^\infty e^{-\lambda(b_2+n+c_2)} \lambda^{(a_2+\sum_{i=1}^{n} y_i)-1} d\lambda$$

Hence from equation (4.12.2) we get

$$p_2(\lambda \mid Y) = \frac{(n+b_2+c_2)^{a_2+\sum_{i=1}^{n} y_i}}{\Gamma\left(a_2+\sum_{i=1}^{n} y_i\right)} e^{-\lambda(b_2+n+c_2)} \lambda^{(a_2+\sum_{i=1}^{n} y_i)-1} \qquad (4.12.3)$$

This is the probability density function of Gamma distribution with parameter $\alpha_2 = (n+b_2+c_2),\, and\, \beta_2 = \left(a_2+\sum_{i=1}^{n} y_i\right)$

4.13 Chi-square- exponential distribution as a double prior:

Now it is assumed that the double prior distribution of λ is chi-square distribution with hyper parameter a_3 and the exponential distn with hyper parameter $'c_3'$ which is given below;

$$P_3(\lambda) \propto \lambda^{\frac{a_3}{2}-1} e^{-\frac{\lambda}{2}} e^{-\lambda c_3}, \lambda > 0$$

Now the posterior distribution of λ for the given data 'Y' is:

$$P_3(\lambda \mid Y) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i} \lambda^{\frac{a_3}{2}-1} e^{-\lambda\left(c_3+\frac{1}{2}\right)}, \lambda > 0$$

$$\Rightarrow p_3(\lambda \mid Y) = k\, e^{-\lambda\left(n+c_3+\frac{1}{2}\right)} \lambda^{(\frac{a_3}{2}+\sum_{i=1}^{n} y_i)-1} \qquad (4.13.1)$$

$$where\; k^{-1} = \int_0^\infty e^{-\lambda\left(n+c_3+\frac{1}{2}\right)} \lambda^{(\frac{a_3}{2}+\sum_{i=1}^{n} y_i)-1} d\lambda$$

Then from equation (4.13.1) we get

$$p_3(\lambda \mid Y) = \frac{\left(n + c_3 + \frac{1}{2}\right)^{\frac{a_3}{2} + \sum_{i=1}^{n} y_i}}{\Gamma\left(\frac{a_3}{2} + \sum_{i=1}^{n} y_i\right)} e^{-\lambda\left(n + c_3 + \frac{1}{2}\right)} \lambda^{(\frac{a_3}{2} + \sum_{i=1}^{n} y_i) - 1} \tag{4.13.2}$$

which is the pdf of Gamma distribution with parameters $\alpha_3 = \left(n + c_3 + \frac{1}{2}\right)$, $and\,\beta_3 = \left(\frac{a_3}{2} + \sum_{i=1}^{n} y_i\right)$ so the posterior distribution of λ for given data is Gamma distribution having parameters α_3 and β_3.

4.14 Gamma distribution as prior:

The single prior distribution of λ is Gamma distribution with hyper parameters a_4 and b_4 which is given below:

$$P_4(\lambda) = \frac{b_4 e^{-\lambda b_4} \lambda^{a_4 - 1}}{\Gamma a_4}, \lambda > 0; a_4, b_4 > 0$$

Now the posterior distribution of λ for the given data 'Y' is

$$P_4(\lambda \mid Y) \propto e^{-\lambda b_4} \lambda^{a_4 - 1} e^{-\lambda n} \lambda^{\sum_{i=1}^{n} y_i}, \lambda > 0$$

$$\Rightarrow P_4(\lambda \mid Y) = k\, e^{-\lambda(b_4 + n)} \lambda^{\left(a_4 + \sum_{i=1}^{n} y_i\right) - 1} \tag{4.14.1}$$

Where $k^{-1} = \int_0^\infty e^{-\lambda(b_4 + n)} \lambda^{\left(a_4 + \sum_{i=1}^{n} y_i\right) - 1} d\lambda$

From equation (4.14.1) we get

$$P_4(\lambda \mid Y) = \frac{(b_4 + n)^{a_4 + \sum_{i=1}^{n} y_i}}{\Gamma\left(a_4 + \sum_{i=1}^{n} y_i\right)} e^{-\lambda(b_4 + n)} \lambda^{\left(a_4 + \sum_{i=1}^{n} y_i\right) - 1} \tag{4.14.2}$$

This is the Gamma distribution with parameters $\alpha_4 = b_4 + n,\ and\,\beta_4 = \left(a_4 + \sum_{i=1}^{n} y_i\right)$.

So the posterior distribution of λ for given data is Gamma distribution having parameter α_4 and β_4. The Gamma distribution is a natural conjugate prior for λ of Poisson distribution (see Gelman et.al (1995) and Bernardo and smith (1994).

4.15 Comparison of priors with respect to posterior variances:

The variances of the posterior distribution under all of assumed informative priors are calculated by assuming different sets of values hyper parameter, which are given in tables 4.3.1, 4.3.2, 4.3.3 the variance of posterior distribution under all assumed priors is

$$v(\lambda \mid Y) = \frac{\beta_i}{\alpha^2{}_i}, i = 1,2,3,4$$

4.16 The posterior predictive distribution:

Since we have observed that there is only type of posterior distribution derived under all the priors i.e. Gamma distribution. We now derive predictive distribution under this posterior distribution.

4.17 Posterior predictive distribution under Gamma chi-square prior:

The posterior predictive distribution for $X = Y_{n+1}$ given $(Y : y_1, y_{2,}......, y_n)$ under Gamma-chi-square distribution is

$$P_1(x \mid Y) = \int_0^\infty P(x \mid \lambda) P_1(\lambda \mid Y) d\lambda \qquad (4.17.1)$$

$$\Rightarrow P_1(x \mid Y) = \int_0^\infty \frac{\alpha_1{}^{\beta_1} e^{-\alpha_1 \lambda} \lambda^{\beta_1 - 1} e^{-\lambda} \lambda^x}{\Gamma\beta_1 \, x!} d\lambda \qquad (4.17.2)$$

$$\Rightarrow P_1(x \mid Y) = \frac{\alpha_1{}^{\beta_1}}{\Gamma\beta_1 \, x!} \int_0^\infty e^{-\lambda(\alpha_1 + 1)} \lambda^{\beta_1 + x - 1} d\lambda$$

$$\Rightarrow P_1(x \mid Y) = \frac{\alpha_1{}^{\beta_1}}{\Gamma\beta_1 \, x!} \frac{\Gamma(\beta_1 + x)}{(\alpha_1 + 1)^{\beta_1 + x}}, x = 0,1,2,......$$

(4.17.3)

This is probability mass function of Poisson-Gamma distribution i.e

$$X \mid Y \sim PG(\beta_1, \alpha_1, 1) \qquad \alpha_1, \beta_1 > 0; x = 0,1,2.....$$

Where α_1 and β_1 are given in (4.11.4)

4.18 The posterior predictive distribution Gamma-exponential prior:

The posterior predictive distribution for $X = Y_{n+1}$ given $(Y : y_1, y_2,..., y_n)$ under Gamma exponential prior:

$$P_2(x \mid Y) = \int_0^\infty \frac{e^{-\lambda} \lambda^x \alpha_2{}^{\beta_2} e^{-\alpha_2 \lambda} \lambda^{\beta_2 - 1}}{\Gamma\beta_2 \, x!} d\lambda$$

$$P_2(x \mid Y) = \frac{\alpha_2{}^{\beta_2}}{\Gamma\beta_2 \, x!} \frac{\Gamma(\beta_2 + x)}{(\alpha_2 + 1)^{\beta_2 + x}}, x = 0,1,2,...... \qquad (4.18.1)$$

This is the probability mass function of Poisson-Gamma distribution i.e

$$X \mid Y \sim PG(\beta_2, \alpha_2, 1) \qquad \alpha_2, \beta_2 > 0; x = 0,1,2.....$$

Where α_2 and β_2 are given in (4.12.3)

4.19 Posterior predictive distribution under chi-square exponential prior:

The posterior predictive distribution of $X = Y_{n+1}$ given$(Y: y_1, y_2,..., y_n)$ under chi-square exponential prior:

$$P_3(x \mid Y) = \int_0^\infty \frac{e^{-\lambda}\lambda^x {\alpha_3}^{\beta_3} e^{-\alpha_3\lambda}\lambda^{\beta_3-1}}{\Gamma\beta_3 \; x!} d\lambda$$

$$P_3(x \mid Y) = \frac{{\alpha_3}^{\beta_3}}{\Gamma\beta_3 \; x!} \frac{\Gamma(\beta_3 + x)}{(\alpha_3 + 1)^{\beta_3 + x}}, x = 0,1,2,...... \qquad (4.19.1)$$

This is also the probability density function of Poisson-Gamma distribution i.e.

$$X \mid Y \sim PG(\beta_3, \alpha_3, 1) \qquad \alpha_3, \beta_3 > 0; x = 0,1,2.....$$

where α_3 andβ_3 are given in (4.13.2)

4.20 Posterior predictive distribution under Gamma prior:

Now, we consider the posterior predictive distribution of $X = Y_{n+1}$ given $(Y: y_1, y_{2,} ......, y_n)$ under Gamma prior is

$$P_4(x \mid Y) = \int_0^\infty \frac{e^{-\lambda}\lambda^x {\alpha_4}^{\beta_4} e^{-\alpha_4\lambda}\lambda^{\beta_4-1}}{\Gamma\beta_4 \; x!} d\lambda$$

$$P_4(x \mid Y) = \frac{{\alpha_4}^{\beta_4}}{\Gamma\beta_4 \; x!} \frac{\Gamma(\beta_4 + x)}{(\alpha_4 + 1)^{\beta_4 + x}}, x = 0,1,2,...... \qquad (4.20.1)$$

which is the pmf of Poisson-Gamma distribution i.e.

$$X \mid Y \sim PG(\beta_4, \alpha_4, 1) \qquad \alpha_4, \beta_4 > 0; x = 0,1,2.....$$

where α_4 and $\beta_4$are given (4.14.2)

4.21 Comparison of prior using the posterior predictive distribution variances:

The posterior predictive variances using different prior distribution are given in the tables 4.3.4; 4.3.2; and 4.3.3

For the posterior predictive (Poisson-Gamma) distribution we have

$$\text{var}(x \mid Y) = \frac{\beta_i}{\alpha_i}\left(1 + \frac{1}{\alpha_i}\right) \quad for \;\; i = 1,2,3,4$$

4.22 Normal Approximation of Possion Distribution:

Suppose$\underline{Y} = (y_1, y_2, ..., y_n)$ is a random sample from Poisson distribution with unknown parameter λ, then $\mathrm{f}(\mathrm{Y} \mid \lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{y}}}{\mathrm{y}!}$.

The likelihood function is given by

$$i.e. \quad L(\lambda \mid Y) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

We consider a more general class of priors, $P(\lambda) \propto \frac{1}{\lambda^{c/2}}, c \geq 0$

To construct the approximation we need the second derivatives of the log-posterior density,

$$\log P(\lambda \mid Y) = \text{constant} - n\lambda + \left(\sum_{i=1}^{n} y_i - \frac{c}{2}\right)\log \lambda$$

The first derivative is

$$\frac{\partial \log P(\lambda \mid Y)}{\partial \lambda} = -n + \frac{1}{\lambda}\left(\sum_{i=1}^{n} y_i - \frac{c}{2}\right)$$

From which the posterior mode is readily obtained as

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

The second derivative of the log posterior density is

$$\frac{\partial^2 \log P(\lambda \mid Y)}{\partial \lambda^2} = -\frac{1}{\lambda^2}\left(\frac{2\sum_{i=1}^{n} y_i - c}{2}\right)$$

and hence negative of the hessian is

$$I(\lambda) = \frac{-\partial \log P(\lambda \mid Y)}{\partial \lambda^2} = \frac{1}{\lambda^2}\left(\frac{2\sum_{i=1}^{n} y_i - c}{2}\right)$$

and therefore $I(\hat{\lambda}) = \left(\frac{2n^2}{2\sum_{i=1}^{n} y_i - c}\right)$

Therefore, the large sample approximate posterior distribution is

$$p(\lambda \mid y) \approx N\left(\frac{2\sum_{i=1}^{n} y_i - c}{2n}, \frac{2\sum_{i=1}^{n} y_i - c}{2n^2}\right) \tag{4.22.1}$$

For c=0, $P(\lambda) \propto 1$ (uniform prior), we have from (4.22.1)

$$p(\lambda \mid y) \approx N\left(\bar{y}, \frac{\bar{y}}{n}\right)$$

For c=1, $P(\lambda) \propto \frac{1}{\lambda^{1/2}}$ (jefferys' prior), we have from (4.22.1)

$$p(\lambda \mid y) \approx N\left(\frac{2\sum_{i=1}^{n} y_i - 1}{2n}, \frac{2\sum_{i=1}^{n} y_i - 1}{2n^2}\right)$$

For c=2, $P(\lambda) \propto \frac{1}{\lambda}$, we have from (4.22.1)

$$p(\lambda \mid y) \approx N\left(\frac{\sum_{i=1}^{n} y_i - 1}{n}, \frac{\sum_{i=1}^{n} y_i - 1}{n^2}\right)$$

Now consider another class of prior (Gamma prior) given by

$$P(\lambda) \propto e^{-\lambda a}\lambda^{b-1}; \lambda > 0' a, b > 0$$

The log posterior density of λ is given by

$$\log P(\lambda \mid Y) = \text{constant} - \lambda(a+n) + \left(b + \sum_{i=1}^{n} y_i - 1\right)\log\lambda$$

The first derivative is $\frac{\partial \log P(\lambda \mid Y)}{\partial\lambda} = -(a+n) + \left(b + \sum_{i=1}^{n} y_i - 1\right)\frac{1}{\lambda}$

From which the posterior mode is readily obtained as

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i + b - 1}{a+n}$$

The second derivative of the log posterior density is

$$\frac{\partial^2 \log P(\lambda \mid Y)}{\partial\lambda^2} = -\frac{1}{\lambda^2}\left(\sum_{i=1}^{n} y_i + b - 1\right)$$

And hence, negative of the hessian is

$$I(\lambda) = \frac{-\partial \log P(\lambda \mid Y)}{\partial\lambda^2} = \frac{1}{\lambda^2}\left(\sum_{i=1}^{n} y_i + b - 1\right)$$

The second derivative at the mode $\hat{\lambda}$ is then $I(\hat{\lambda}) = \left(\dfrac{(a+n)^2}{\sum_{i=1}^{n} y_i + b - 1} \right)$

Therefore, the large sample approximate posterior distribution is

$$p(\lambda \mid y) \approx N\left(\frac{\sum_{i=1}^{n} y_i + b - 1}{a+n}, \frac{\sum_{i=1}^{n} y_i + b - 1}{(a+n)^2} \right)$$

Thus 95% approximate HPD credible interval for λ when general class of priors $P(\lambda) \propto \dfrac{1}{\lambda^{c/2}}$ is considered

$$\left[\frac{2\sum_{i=1}^{n} y_i - c}{2n} - 1.96\sqrt{\frac{2\sum_{i=1}^{n} y_i - c}{2n}},\ \frac{2\sum_{i=1}^{n} y_i - c}{2n} + 1.96\sqrt{\frac{2\sum_{i=1}^{n} y_i - c}{2n}} \right]$$

4.23 Laplace Approximations For Poisson Distribution:

The probability density function of Poisson distribution is

$$f(Y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0,1,2,........: \lambda > 0$$

The likelihood function of Poisson distribution is

$$L(\lambda \mid Y) = \frac{e^{-n\lambda} \ \lambda^{n\bar{y}}}{\prod_{i=1}^{n} y!}$$

Consider gamma as priors

$$P(\lambda, \alpha, \beta) = \frac{\beta^{\alpha} e^{-\beta\lambda} \lambda^{\alpha-1}}{\Gamma\alpha}, c > 0$$

Thus Posterior $\propto$ likelihood x prior

$$P(\lambda \mid Y) \propto \frac{e^{-n\lambda}\lambda^{n\bar{y}}}{\prod_{i}^{n} y_i} \frac{\beta^{\alpha}}{\Gamma\alpha} e^{-\beta\lambda}\lambda^{\alpha-1}$$

$$\Rightarrow P(\lambda \mid Y) = cons\tan t \ e^{-\lambda(\lambda+\beta)}\lambda^{n\bar{y}+\alpha-1}$$

To construct the Laplace approximation, we need posterior mode $\hat{\lambda}$ and $I(\hat{\lambda})$ of the log-posterior density.

$$\log P(\lambda \mid Y) = constant - \lambda(y+\beta) + (n\bar{y}+\alpha-1)\log\lambda$$

The posterior mode of this density is readily obtained as

$$\frac{\partial \log P(\lambda \mid Y)}{\partial \lambda} = -(n+\beta) + \frac{(n\bar{y}+\alpha-1)}{\lambda} = 0$$

$$\Rightarrow \hat{\lambda} = \frac{\mathrm{n}\bar{\mathrm{y}}+\alpha-1}{\mathrm{n}+\beta}$$

For $\alpha = 1, \& \beta = 0$; $\hat{\lambda} = \bar{y}$ which is same as MLE of Poisson distribution

The second derivative of the log-posterior density at mode $\hat{\theta}$ is

$$\frac{\partial^2 \log P(\lambda \mid Y)}{\partial \lambda^2} = \frac{-(n\bar{y}+\alpha-1)}{\lambda^2}$$

$$\Rightarrow \frac{\partial^2 \log P(\lambda \mid Y)}{\partial \lambda^2} = \frac{-(n+\beta)^2}{n\bar{y}+\alpha-1}$$

$$\Rightarrow \left|I(\hat{\lambda})\right| = \frac{(n+\beta)^2}{n\bar{y}+\alpha-1}$$

The Laplace's approximation to the posterior of Poisson distribution is given by

$$P(\lambda \mid Y) \approx (2\pi)^{\frac{-1}{2}} \left|I(\hat{\lambda})\right|^{-\frac{1}{2}} \exp\left[\log P(\lambda \mid Y) - \log P(\hat{\lambda} \mid Y)\right]$$

$$P(\lambda \mid Y) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+\beta)}{(n\bar{y}+\alpha-1)^{1/2}} \exp\left[-\lambda(n+\beta) + (n\bar{y}+\alpha-1)\log\lambda + \hat{\lambda}(n+\beta) - (n\bar{y}+\alpha-1)\log\hat{\lambda}\right]$$

$$P(\lambda \mid Y) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+\beta)}{(n\bar{y}+\alpha-1)^{1/2}} \exp\left[-\lambda(n+\beta) + (n\bar{y}+\alpha-1)\log\lambda + \frac{(n\bar{y}+\alpha-1)}{(n+\beta)}(n+\beta) - (n\bar{y}+\alpha-1)\log\left(\frac{n\bar{y}+\alpha-1}{n+\beta}\right)\right]$$

$$P(\lambda \mid Y) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+\beta)}{(n\bar{y}+\alpha-1)^{1/2}} \exp\left[-\lambda(n+\beta) + (n\bar{y}+\alpha-1)\right] \exp\left[\log\left(\frac{n+\beta}{(n\bar{y}+\alpha-1)}\right)^{n\bar{y}+\alpha-1}\right]$$

4.24 Lindely Approximation for Possion Distribution:

Suppose $Y = (y_1, y_2, \ldots\ldots, y_n)$ is a random sample from Poisson distribution with unknown parameter λ then the pmf of Poisson distribution is

$$f(Y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad y = 0,1,2,\ldots..; \lambda > 0$$

And the likelihood function is

$$L(\lambda \mid Y) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!},$$

We consider here that the prior distribution of λ is lognormal with density

$$P(\theta)=\frac{1}{\lambda\sqrt{2\pi}}\exp\left[\frac{-1}{2\phi}(\log\lambda-\mu)^2\right];\ \lambda>0$$

Where (μ,ϕ) are known hyper parameter.

Since $L(\lambda)=-n\lambda+\sum_{i=1}^{n}Y_i\log\lambda+cons\tan t$

$$\Rightarrow L_1(\lambda)=\frac{\partial L(\lambda)}{\partial\lambda}=0 \text{ gives } \hat{\lambda}=\bar{Y}$$

$$\Rightarrow L_2(\lambda)|_{\lambda=\hat{\lambda}}=\frac{-\sum_{i=1}^{n}Y_i}{\hat{\lambda}^2}=\frac{-n\bar{y}}{\bar{y}^2}=\frac{-n}{\bar{y}}$$

$$\Rightarrow L_3(\lambda)|_{\lambda=\hat{\lambda}}=\frac{-2\sum_{i=1}^{n}Y_i}{\hat{\lambda}^3}=\frac{2n\bar{y}}{\bar{y}^3}=\frac{2n}{\bar{y}^2}$$

$$\Rightarrow \sigma^2|_{\lambda=\hat{\lambda}}=\left[\frac{\sum_{i=1}^{n}Y_i}{\hat{\lambda}^2}\right]^{-1}=\frac{\bar{y}}{n}$$

Further

$$\rho(\lambda)=\frac{-1}{2\phi}(\log\lambda-\mu)^2-\log\lambda+cons\tan t$$

$$\Rightarrow \rho_1(\lambda)=\frac{\partial}{\partial\lambda}\rho(\lambda)$$

$$\Rightarrow \rho_1(\lambda)|_{\lambda=\hat{\lambda}}=\frac{-1}{\bar{y}}\left[1+\frac{\log\bar{y}-\mu}{\phi}\right]$$

Hence, $E\left(\lambda\,|\,\underset{\sim}{Y}\right)\cong\bar{y}-\frac{1}{\bar{y}}\left[1+\frac{\log\bar{y}-\mu}{\phi}\right]\frac{\bar{y}}{n}+\frac{n}{\bar{y}^2}\frac{\bar{y}^2}{n^2}$

$$E\left(\lambda\,|\,\underset{\sim}{Y}\right)\cong\bar{y}-\frac{1}{n}\left[1+\frac{\log\bar{y}-\mu}{\phi}\right]+O\left(\frac{1}{n^2}\right)$$

4.25 Bayesian Estimation of Prior Distribution Function of the Possion Model:

The distribution function of a Poisson distribution with parameter λ obtained by repeated integration by parts is given by

$$F(\lambda,t)=\sum_{y=0}^{t}\frac{e^{-\lambda}\lambda^y}{y!}=1-\frac{Y(t+1,\lambda)}{\Gamma(t+1)} \qquad (4.25.1)$$

Where $Y(a,b)=\int_0^b e^{-y}\, y^{a-1}\, dy,$ the incomplete gamma function and $\Gamma a=\int_0^\infty e^{-y}\, y^{a-1}\, dy$, the complete gamma function. The relationship (4.25.1) between the two prominent distributions, one discrete and other continuous turns out to be significant in many studies, especially in the study of stochastic point process. In certain situations, the Poisson distribution gives a very good approximation to the binomial distribution.

The gamma distribution is the natural conjugate prior for λ of a Poisson model (Gelman et al, 1995). A family of priors is the conjugate if the choice of a prior in that family generates a posterior that belongs to a same family. We here consider a gamma prior, $G(\alpha,\beta)$ as describing the prior uncertainty about the parameter λ having the pdf

$$P(\lambda)=\frac{\alpha^{\beta}}{\Gamma\beta}e^{-\alpha\lambda}\,\lambda^{\beta-1},\quad \alpha,\beta,\lambda>0 \tag{4.25.2}$$

where α and β are the prior parameters. The parameter α is the shape parameter while β is a scale parameter.

We now will derive Bayes estimator of $F(\lambda,t)$ under the prior (4.25.2) and also report the MLE of $F(\lambda,t)$. The likelihood function for a random sample $Y=(y_1,y_2,\ldots\ldots,y_n)$ of size n from a Poisson distribution is given by

$$L(\lambda\,|\,Y)\propto e^{-n\lambda}\,\lambda^{T},\ where T=\sum_{i=1}^{n} y_i \tag{4.25.3}$$

Combining the likelihood function (4.25.3) and the prior (4.25.2) the posterior of λ is

$$P(\lambda\,|\,Y)=\frac{(\alpha+\beta)^{\beta+\sum_{i=1}^{n}y_i}}{\Gamma\left(\beta+\sum_{i=1}^{n}y_i\right)}e^{-\lambda(\alpha+n)}\lambda^{\left(\beta+\sum_{i=1}^{n}y_i\right)-1}$$

$$\Rightarrow P(\lambda\,|\,Y)=\frac{(\alpha')^{\beta'}\,e^{-\lambda\alpha'}\,\lambda^{\beta'-1}}{\Gamma\beta'},\lambda>0 \tag{4.25.4}$$

where $\alpha'=\alpha+n$ and $\beta'=\beta+T$. The posterior pdf (4.25.4) is also gamma, $G(\alpha',\beta')$, showing that the posterior distribution has the same functional from as the prior, and, hence the gamma priors are closed under sampling.

The Bayes estimator of $F(\lambda,t)$ under the squared error loss function is

$$F^*(\lambda,t) = E[F(\lambda,t) \mid Y] = 1 - \int_0^\infty \frac{Y(t+1,\lambda)}{\Gamma(t+1)} \frac{(\alpha')^{\beta'} e^{-\lambda\alpha'} \lambda^{\beta'-1}}{\Gamma\beta'} d\lambda \ , \lambda > 0$$

$$= 1 - \frac{(\alpha')^{\beta'}}{\Gamma\beta'\Gamma(t+1)} \int_0^\infty e^{-\lambda\alpha'} + \lambda^{\beta'-1}\gamma(t+1.\lambda)d\lambda \qquad (4.25.5)$$

In order to evaluate (4.25.5), we use result on the Laplace transformation of the incomplete gamma function,(Erdelyi,1953)

$$\int_0^\infty e^{-ax} x^{p-1} \gamma(q,x)dx = \frac{\Gamma(p+q)}{q(1+a)^{p+q}} \ {}_2F_1\left[1, p+q; q+1; \frac{1}{1+a}\right] \qquad (4.25.6)$$

Subject to $R_e\, a > 0, R_e\,(p+q) > 0,$ where ${}_2F_1\left[1, p+q; q+1; \frac{1}{1+a}\right]$ is the Gauss hyper geometric function. Setting $a = \alpha', p = \beta', q = t+1$ in (4.25.6), the Bayes estimator, $F^*(\lambda, t)$ of $F(\lambda, t)$ in (4.25.5) takes the form:

$$F^*(\lambda,t) = 1 - \frac{(\alpha')^{\beta'}\Gamma(\beta'+t+1)}{(t+1)\Gamma(\beta')\Gamma(t+1)(1+\alpha')^{\beta'+t+1}} \ {}_2F_1\left[1, \beta'+t+1; t+2; (1+\alpha')^{-1}\right]$$

$$\Rightarrow F^*(\lambda,t) = 1 - \frac{(\alpha')^{\beta'}(1+\alpha')^{-(\beta'+t+1)}}{(t+1)B(t+1,\beta')} {}_2F_1\left[1, \beta'+t+1; t+2; (1+\alpha')^{-1}\right] \qquad (4.25.7)$$

where $B(t+1,\beta') = \frac{\Gamma(t+1)\Gamma(\beta')}{\Gamma(t+1+\beta')}$, the beta function.

To evaluate ${}_2F_1\left[1, \alpha'+t+1; t+2; (\beta'+1)^{-1}\right]$ in (4.25.7) we the following relationship between the Gauss hyper geometric function and the gamma function Abramowitz stegun (1972),

$${}_2F_1[a,b;c;z] = \frac{\Gamma c}{\Gamma a \Gamma b} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}$$

(4.25.8)

It is well-known that the MLE λ of a Poisson distribution is $\bar{y}$ and hence by the invariance property, the MLE of $F(\lambda, t)$ in (4.25.1) is

$$\hat{F}(\lambda,t) = 1 - \frac{Y(t+1,\bar{y})}{\Gamma(t+1)} \qquad (4.25.9)$$

4.26 Estimation of the left truncated $F(\lambda, t)$: zero class missing:

When $y < c$ in a Poisson distribution cannot be observed or are missing and the remaining probabilities at $y = c, c+1, \ldots\ldots, \infty$ adjusted so that $\sum_c^\infty f(y) = 1$, is called the left truncated Poisson distribution below c. The probability function of the Poisson distribution truncated at $y = 0$ is given by

$$f(y) = \frac{e^{-\lambda}\lambda^{y}}{\left(1-e^{-\lambda}\right)y!}, \quad y = 0,1,2,.....;\lambda > 0 \tag{4.26.1}$$

Using the negative binomial expansion, namely,

$$\left(1-e^{-\lambda}\right)^{n} = \sum_{j=0}^{\infty} \binom{n+j-1}{j} e^{-\lambda j}$$

The likelihood function can be written as

$$L(Y|\lambda) \propto \lambda^{T} \sum_{j=0}^{\infty} \binom{n+j-1}{j} e^{-\lambda(j+1)} \tag{4.26.2}$$

where $T = \sum_{i=1}^{n} y_i$

Combining the prior (4.25.2) and the likelihood function (4.26.2), the posterior pdf of λ is,

$$P(\lambda|Y) = k\sum_{j=0}^{\infty} \binom{n+j-1}{j} \lambda^{\beta'-1} e^{-\lambda(\alpha'+j)} \tag{4.26.3}$$

where $\alpha' = \alpha + \mathrm{n}$ and $\beta' = \beta + \mathrm{T}$ and $k^{-1} = \sum_{j=0}^{\infty} \binom{n+j-1}{j} \frac{\Gamma\beta'}{(\alpha'+j)^{\beta'}}$ (4.26.4)

The Bayes estimator of λ under the squared error loss function is

$$\lambda^{*} = E(\lambda|Y) = k\sum_{j=0}^{\infty} \binom{n+j-1}{j} \int_{0}^{\infty} \lambda^{\beta} e^{-\lambda(\beta'+j)} d\lambda$$

$$\Rightarrow \lambda^{*} = \frac{\sum_{j=0}^{\infty} \binom{n+j-1}{j} \Gamma(\beta'+1)}{\sum_{j=0}^{\infty} \binom{n+j-1}{j} \frac{\Gamma\beta'}{(\alpha'+j)^{\beta'}} (\beta'+j)^{\beta'+1}}$$

$$\Rightarrow \lambda^{*} = \frac{\beta' \sum_{j=0}^{\infty} \binom{n+j-1}{j} (\alpha'+j)^{-(\beta'+1)}}{\sum_{j=0}^{\infty} \binom{n+j-1}{j} (\alpha'+j)^{-(\beta')}} \tag{4.26.5}$$

The distribution function $\mathrm{F}_0(\lambda, \mathrm{t})$ with zero class missing is given by

$$\mathrm{F}_0(\lambda,\mathrm{t}) = \left(1-\mathrm{e}^{-\lambda}\right)^{-1}\left[1-\mathrm{e}^{-\lambda} - \frac{\mathrm{y}(\mathrm{t}+1,\lambda)}{\Gamma(\mathrm{t}+1)}\right] \tag{4.26.6}$$

Thus the Bayes estimator of $\mathrm{F}_0(\lambda, \mathrm{t})$ under squared error loss function is

$$F_0 *(\lambda, t) = E[F_0(\lambda, t) \mid Y] = k\int_0^\infty \left(1-e^{-\lambda}\right)^{-1}\left[1-e^{-\lambda} - \frac{y(t+1,\lambda)}{\Gamma(t+1)}\right]\sum_{j=0}^{\infty}\binom{n+j-1}{j}\lambda^{\beta'-1}e^{-\lambda(\alpha'+1)}d\lambda \quad (4.26.7)$$

where k is given by (4.26.4). Using $\left(1-e^{-\lambda}\right)^{-1}$, the equation (4.26.7) takes the form

$$F^*{}_0(\lambda, t) = k\sum_{j=0}^{\infty}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\int_0^\infty \lambda^{\beta'-1}e^{-\lambda(\alpha'+j+i)}d\lambda$$

$$= k\sum_{j=0}^{\infty}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\int_0^\infty \lambda^{\beta'-1}e^{-\lambda(\alpha'+j+i)}d\lambda - k\sum_{j=0}^{\infty}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\int_0^\infty \lambda^{\beta'-1}e^{-\lambda(j+i+1+\alpha')}d\lambda - k\sum_{j=0}^{\infty}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\int_0^\infty \lambda^{\beta'-1}e^{-\lambda(\alpha'+j+i)}\frac{\gamma(t+1,\lambda)}{\Gamma(t+1)}d\lambda \quad (4.26.8)$$

The first integral within the double summation of (4.26.8) is a complete gamma function which can easily be seen as $\frac{\Gamma\beta'}{\left(\alpha' + j + i\right)^{\beta'}}$. Similarly the second integral becomes $\frac{\Gamma\beta'}{\left(\alpha' + j + i + 1\right)^{\beta'}}$. The third integral within the double summation can be evaluated by using the Laplace transformation of the incomplete gamma function as in (4.7.6) by setting $a = \alpha' + j + i, p = \beta', q = t + 1$. Hence the third integral takes the form

$$\frac{\Gamma\left(\beta' + t + 1\right)}{(t+1)!\left(\alpha' + j + i + 1\right)^{\beta' + t + 1}} \; {}_2F_1\left[1, \beta' + t + 1; t + 2; \left(\alpha' + j + i + 1\right)^{-1}\right] \quad (4.26.9)$$

Finally, the equation (4.26.8) becomes

$$F^*{}_0(\lambda, t) = k\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\binom{n+j-1}{j}\left\{\left[\left(\alpha' + j + i\right)^{-\beta'} - \left(\alpha' + j + i + 1\right)^{-\beta'}\right]\Gamma\beta' - \frac{\Gamma\left(\beta' + t + 1\right)}{(t+1)!\left(\alpha' + j + i + 1\right)^{\beta' + t + 1}} \; {}_2F_1\left[1, \beta' + t + 1; t + 2; \left(\alpha' + j + i + 1\right)^{-1}\right]\right\} \quad (4.26.10)$$

We note that the mean of the possion distribution truncated at y=0 is $\frac{\lambda}{\left(1-e^{-\lambda}\right)}$. Hence the MLE of the λ for the given truncated sample can be obtained by solving the equation numerically;

$$\frac{\hat{\lambda}}{\left(1-e^{-\hat{\lambda}}\right)} = \bar{y}^* \quad (4.26.11)$$

where $\bar{y}^*$ is the truncated sample mean (Irwin, 1959) has given an explicit expression for $\hat{\lambda}$ of (4.26.11) as

$$\hat{\lambda} = \bar{y}^* - \sum_{j=1}^{\infty}\frac{j^{j-1}}{j!}\left(\bar{y}^* e^{-\bar{y}^*}\right)^j \quad (4.26.12)$$

Again by the invariance property, the MLE of $F_0(\lambda, t)$ in (4.26.6) is

$$\hat{F}_0(\lambda, t) = \left(1 - e^{-\bar{y}_0^*}\right)^{-1}\left[1 - e^{-\bar{y}^*_0} - y\left(t+1; \bar{y}_0^*\right)/\Gamma(t+1)\right] \quad (4.26.13)$$

where $\bar{y}_0^*$ is the solution of equation (4.26.6) for $\hat{\lambda}$.

Example 4.1(Gelman et al, 1995): Suppose we have information on deaths of passengers in the airline accidents from 1976 to 1985

Year:	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Passenger deaths:	734	516	754	877	814	362	764	809	223	1066

We have developed some programmes for MLLE and Bayes estimates for the above example in R software

<u>#Program for MLE of Poisson distribution.</u>

```
library(stats4)
post.pois<-function(theta,y){
return(length(y)*theta-sum(y)*log(theta)-sum(log(y)))
}
y<-c(734,516,754,877,814,362,764,809,223,1066)
opt <- optim(par = 552, fn = post.pois, method =
"BFGS",y=y)
$par
[1] 691.9003
$value
[1] -38391.95
$counts
function gradient
       8        6
$convergence
[1] 0
$message
NULL
```

 # Bayes estimates of Poisson distribution.

```
bayes.est<-function(y)
 {
 n<-length(y)
 C<-c(0,1)
 estimate<-(sum(y)-(C/2)+1)/n
```

```
 return(estimate)
 }
y<-c(734,516,754,877,814,362,764,809,223,1066)
bayes.est(y=y) # To get the output.
 [1] 692.00 691.95
```

Example 4.2 (Birth rates):

The Poisson distribution provides a realistic model for many random phenomena. Since the values of a Poisson random variable are non negative integers, any random phenomena for which a count of some sort is of interest, is a candidate for modeling by assuming a Poisson distribution.

Over a course of the 1990s the General Social Survey gathered data on the educational attainment and number of children of 155 women who were 40 years of age at the time of their participation in the survey. Let $Y_{1,1}, Y_{2,1}, ..., Y_{n1,1}$ denote the number of children for n_1 women without college degrees and $Y_{1,2}, Y_{2,2}, ..., Y_{n2,2}$ denote the number of children for n_2 women with college degrees. The group sums and means are as follows:

Less than bachelor's: $n_1 = 111, \quad \sum_{i=1}^{n_1} Y_{i,1} = 217, \quad \bar{Y}_1 = 1.95$

Bachelors or higher: $n_2 = 44, \quad \sum_{i=1}^{n_2} Y_{i,2} = 66, \quad \bar{Y}_1 = 1.50$

Posterior means, variances are obtained from their gamma posterior distributions by using the following program in R software and are presented in table 4.2.1.

 # Posterior mean and variance under gamma chi-square as a double prior

```
Post.mav<-function(a1,b1,c1,n,sy){
pmgc<-(a1+c1/2+sy-1)/(n+b1+1/2)
pvgc<-(a1+c1/2+sy-1)/((n+b1+1/2)^2)
list(Posterior.mean=pmgc,Posterior.variance=pvgc)
 }
 Post.mav(2,2,2,111,217)
```

 # Posterior mean and variance under gamma exponential as a double prior

```
 Post.mav<-function(a2,b2,c2){
 n<-111;sy1<-217
 pmge<-(a2+sy1)/(n1+b2+c2)
```

```
pvge<-(a2+sy1)/((n1+b2+c2)^2)
 list(Posterior.mean=pmge,Posterior.variance=pvge)
 }
 Post.mav(2,2,2)
```

 # Posterior mean and variance under chi-square exponential as a double prior

```
 Post.mav<-function(a3,b3,c3){
n<-111;sy1<-217
pmce<-(a3/2+sy1)/(n1+c3+1/2)
pvce<-(a3/2+sy1)/((n1+c3+1/2)^2)
list(Posterior.mean=pmce,Posterior.variance=pvce)
 }
Post.mav(2,2,2)
```

 # Posterior mean and variance under gamma prior

```
 Post.mav<-function(a,b){
 n<-111;sy1<-217
pmg<-(b+sy1)/(a+n1)
pvg<-(b+sy1)/((a+n1)^2)
list(Posterior.mean=pmg,Posterior.variance=pvg)
 }
 Post.mav(2,2,2)
```

Table 4.2.1:

Type of Prior	**Less than bachelor's**		**Bachelors or higher**	
	Posterior Mean	**Posterior variance**	**Posterior Mean**	**Posterior variance**
Uniform Prior	1.963964	0.01769337	1.522727	0.03460744
Jeffrey's Prior	1.959459	0.01765279	1.511364	0.03434917
Gamma distribution	1.938053	0.01715091	1.478261	0.03213611
Gamma –Chi-square distribution	1.929515	0.01700014	1.462366	0.03144872
Gamma -Exponential distribution	1.904348	0.01655955	1.416667	0.02951389
Chi-square-Exponential distribution.	1.920705	0.01692251	1.440860	0.03098624

We observe that the posterior mean for group1(less than bachelors) under all assumed priors is more than that of group2 (bachelors or higher).The posterior variance of group1(less than bachelors) is less than the posterior variance of group2 (bachelors or higher) under all assumed priors.

Also the posterior variance under all the assumed priors is calculated by assuming the value of hyper parameters to be 2. The posterior variances under the double prior Gamma -Exponential distribution are less as compared to other assumed priors, which shows that this prior is efficient as compared to other priors and this less variation in posterior distribution helps in making more precise Bayesian estimation of the true unknown parameter λ of Poisson distribution.

Example 4.3 (Simulation): We have generated a sample of size 30, 60, 100 from Poisson pmf with parameter λ to represent small, moderate and large sample sizes. Also we have taken different values for parameter λ and hyper parameters.

Programme for simulation in R-software for posterior variance under different Priors:

```
sim.var<-function(y,ai,bi,ci){
n<-length(y)
pvgc<-(ai+ci/2+sum(y)-1)/((n+bi+1/2)^2)
pvge<-(ai+sum(y))/((n+bi+ci)^2)
pvce<-(ai/2+sum(y))/((n+ci+1/2)^2)
pvg<-(bi+sum(y))/((ai+n)^2)
list(pvgc=pvgc,pvge=pvge,pvce=pvce,pvg=pvg)
}
y<-rpois(100,2)
sim.var(y,2,2,2)
```

***# Simulations in R Software for posterior predictive variance of the posterior distribution under different Priors*:**

```
pre.var<-function(y,ai,bi,ci){
n<-length(y)
pvgc<-(ai+ci/2+sum(y)-1)/(n+bi+1/2)*(1+1/( n+bi+1/2))
pvge<-(ai+sum(y))/(n+bi+ci)*(1+1/( n+bi+ci))
pvce<-(ai/2+sum(y))/(n+ci+1/2)*(1+1/( n+ci+1/2))
pvg<-(bi+sum(y))/(ai+n)*(1+1/( ai+n))
list(pvgc=pvgc,pvge=pvge,pvce=pvce,pvg=pvg)
}
```

Table 4.3.1: Variances of the posterior distribution using different priors with n=30.

Size	λ	Hyper Parameters $a_i=b_i=c_i$	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
30	2.0	2	0.06248	0.05709	0.06153	0.06445
		5	0.05594	0.04312	0.05276	0.05632
		8	0.05059	0.03402	0.04587	0.04986
		10	0.04755	0.02960	0.04206	0.04625
	5.0	2	0.13538	0.12370	0.13443	0.13964
		5	0.11704	0.09125	0.11386	0.11918
		8	0.10254	0.07041	0.09782	0.10318
		10	0.09449	0.06040	0.08901	0.09437
	8.0	2	0.23195	0.21193	0.23109	0.23925
		5	0.19797	0.15500	0.19480	0.20244
		8	0.17136	0.11862	0.16663	0.17382
		10	0.15668	0.10120	0.15119	0.15812

Table 4.3.2: Variances of the posterior distribution using different priors with n=60.

Size	λ	Hyper Parameters a_i=b_i=c_i	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
60	2.0	2	0.02944	0.02807	0.02918	0.02991
		5	0.02785	0.02408	0.02692	0.02792
		8	0.02642	0.02094	0.02493	0.02616
		10	0.02555	0.01921	0.02374	0.02510
	5.0	2	0.07756	0.07397	0.07731	0.07882
		5	0.07167	0.06244	0.07074	0.07242
		8	0.06649	0.05349	0.06500	0.06682
		10	0.06337	0.04859	0.06156	0.063469
	8.0	2	0.11929	0.11376	0.11904	0.12122
		5	0.10966	0.09571	0.10873	0.11100
		8	0.10123	0.081717	0.09973	0.10207
		10	0.09617	0.07406	0.09436	0.09673

Table 4.3.3: Variances of the posterior distribution using different priors with n=100.

Size	λ	Hyper Parameters $a_i=b_i=c_i$	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
100	2.0	2	0.01875	0.01821	0.01865	0.01893
		5	0.01810	0.01652	0.01774	0.01814
		8	0.01749	0.01508	0.01690	0.01740
		10	0.01711	0.01423	0.01637	0.01694
	5.0	2	0.04606	0.04474	0.04597	0.04652
		5	0.04388	0.04024	0.04353	0.04417
		8	0.04187	0.03641	0.04128	0.04200
		10	0.04062	0.03416	0.03988	0.04066
	8.0	2	0.07471	0.07257	0.07462	0.07545
		5	0.07093	0.06512	0.07057	0.07147
		8	0.06744	0.05878	0.06685	0.06781
		10	0.06527	0.05506	0.06453	0.06553

The posterior variances under the double prior Gamma- exponential distribution are less as compared to other informative priors, which show that this prior is efficient as compared to the other prior and this less variation in posterior distribution helps in making more priors Bayesian estimation of true unknown parameter λ of Poisson distribution. The results obtained using above programme are presented in tables 4.3.1;4.3.2;4.3.3 for different values of hyper parameters, n and λ.

Table 4.3.4: Posterior Predictive Variances of the posterior distribution using different priors with n=30.

Size	λ	Hyper Parameters a_i=b_i=c_i	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
30	2.0	2	2.09325	1.99827	2.06153	2.12695
		5	2.04185	1.76812	1.92601	2.02775
		8	1.99865	1.59924	1.81211	1.94459
		10	1.97348	1.50960	1.74577	1.89625
	5.0	2	4.53538	4.32958	4.50366	4.60839
		5	4.27197	3.74125	4.15612	4.29061
		8	4.05059	3.30954	3.86405	4.02423
		10	3.92165	3.08040	3.69394	3.86937
	8.0	2	7.77041	7.41782	7.73869	7.89550
		5	7.22614	6.35500	7.11029	7.28816
		8	6.76876	5.57514	6.58222	6.77908
		10	6.50236	5.16120	6.27465	6.48312

Table 4.3.5: Posterior Predictive Variances of the posterior distribution using different priors with n=60.

Size	λ	Hyper Parameters $a_i=b_i=c_i$	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
60	2.0	2	1.86944	1.82495	1.85318	1.88475
		5	1.85228	1.70979	1.79028	1.84331
		8	1.83664	1.61305	1.73296	1.80558
		10	1.82697	1.55671	1.69750	1.78224
	5.0	2	4.90931	4.79248	4.89305	4.94953
		5	4.75083	4.41938	4.68882	4.76449
		8	4.60642	4.10595	4.50274	4.59602
		10	4.51707	3.92343	4.38761	4.49183
	8.0	2	7.57529	7.39502	7.55904	7.63735
		5	7.29287	6.79571	7.23087	7.32639
		8	7.03553	6.29224	6.93185	7.04325
		10	6.87631	5.99906	6.74684	6.86816

Table 4.3.6: Posterior Predictive Variances of the posterior distribution using different priors with n=100.

Size	λ	Hyper Parameters $a_i=b_i=c_i$	Gamma Chi-Square Distribution	Gamma Exponential distribution	Chi-Square Exponential distribution	Gamma distribution
100	2.0	2	1.94070	1.91244	1.93085	1.95030
		5	1.92805	1.83471	1.88978	1.92290
		8	1.91611	1.76508	1.85100	1.89703
		10	1.90851	1.72256	1.82633	1.88057
	5.0	2	5.01430	4.94129	5.00445	5.03912
		5	4.91343	4.69686	4.87516	4.92263
		8	4.81819	4.47792	4.75308	4.81267
		10	4.75760	4.34423	4.67541	4.74272
	8.0	2	7.65444	7.54299	7.64459	7.69233
		5	7.47779	7.15537	7.43952	7.49932
		8	7.31100	6.80819	7.24589	7.31713
		10	7.20488	6.59618	7.12270	7.20124

The results obtained using above programme are presented in tables 4.3.4; 4.3.5; 4.3.6 for different values of hyper parameters, n and λ.

In the tables 4.3.4; 4.3.5; and 4.3.6 It is observed that the values of the posterior predictive variances computed under the double prior Gamma- exponential distribution using different values of hyper parameters are less as compared to the other priors, which means we can prefer the prior Gamma- exponential as a suitable double prior for the unknown parameter λ of Poisson distribution. Further this less variation in the posterior predictive distribution will help us in closely estimating the true probabilities of the future observations.

CHAPTER – 5

BAYESIAN ESTIMATION FOR NORMAL DISTRIBUTION

5.1 Introduction:

The normal distribution plays a very important role in the statistical theory as well as methods. The names of the great mathematician such as Gauss, Laplace, Legendre & others are associated with the discovery & use of the distribution of errors of measurement. The earliest published derivation of the normal distribution was an approximation to a binomial distribution by de-Morvie in 1733. In 1774 Laplace obtained the normal distribution as an approximation to hyper-geometric distribution and advocated tabulation of the probability integral $\Phi(y)$.The work of Gauss in 1809, 1816 respectively established techniques based on the normal distribution which became standard methods used during the nineteenth century. Davis (1952) has shown that the normal distributions give quite a good fit for the failure time data. In 1961 Bazovsky discussed the use of the normal distribution in life testing & reliability problems.

The pdf of the normal distribution with location parameter μ and scale parameter σ in given by

$$f(y \mid \mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \qquad ;-\infty<y<\infty\ ;\ -\infty<\mu<\infty\ ;$$

$\sigma>0$

with mean μ and variance σ^2, $\sigma>0$

5.2 Maximum likelihood estimate of normal distribution:

Let $Y=y_1,y_{2,},\ldots\ldots,y_n$ be a random sample of size n with pdf

$$f(y \mid \mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \qquad ;-\infty<y<\infty\ ;\ -\infty<\mu<\infty$$

Then likelihood is given by

$$L(\mu,\sigma \mid Y)=\prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y_i-\mu)^2}{2\sigma^2}\right\}\right]$$

$$\Rightarrow\ L(\mu,\sigma \mid Y)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\exp\left\{\frac{-\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}$$

The log likelihood is given by

$$LogL(\mu,\sigma\,|\,Y)=\frac{-n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

Case 1: when σ^2 is known, the likelihood equation for estimating μ is:

$$\frac{\partial\log L(\mu,\sigma\,|\,Y)}{\partial\sigma^2}=0$$

$$\Rightarrow\frac{-1}{2\sigma^2}\sum_{i=1}^{n}2(y_i-\mu)(-1)=0$$

$$\Rightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}y_i=\bar{y}$$

Case 2: when μ is known, the likelihood equation for estimating σ^2 is

$$\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma\,|\,Y)=0$$

$$\Rightarrow\ \frac{-n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\mu)^2=0$$

$$\Rightarrow\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2$$

Case 3: Both unknown: The likelihood equation for simultaneous estimation of μ and σ^2 are;

$$\frac{\partial}{\partial\mu}\log L(\mu,\sigma\,|\,Y)=0 \quad \text{and} \quad \frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma\,|\,Y)=0$$

Thus giving $\hat{\mu}=\bar{y}$ and $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu})^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{x})^2=s^2$, sample variance

5.3 Bayesian Estimation for the Parameters of Normal distribution:-

Consider two parameter normal distribution

$$f(y\,|\,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \qquad ;-\infty<y<\infty\ ;\ -\infty<\mu<\infty$$

where μ is the location parameter and σ is the scale parameter. The standard argument as given in Box & Tiao (1973) leads to the quasi prior $P(\mu,\sigma)\propto\frac{1}{\sigma}$, $\sigma>0$ or class of priors $P(\mu,\sigma\,|\,c)\propto\frac{1}{\sigma^{c}}$, $c>0$ which we consider here.

The likelihood function is given by

$$L(\mu,\sigma\,|\,Y)=\prod_{i=1}^{n} f(y_i\,|\,\mu,\sigma)$$

$$\Rightarrow L(\mu,\sigma\,|\,Y)=\frac{1}{\sigma^{n}(2\pi)^{n/2}}\exp\left\{\frac{-1}{2\sigma^{2}}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

$$\Rightarrow L(\mu,\sigma\,|\,Y)=\frac{1}{\sigma^{n}(2\pi)^{n/2}}\exp\left\{\frac{-1}{2\sigma^{2}}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2\right]\right\}$$

$\Rightarrow$

$$L(\mu,\sigma\,|\,Y)=\frac{1}{\sigma^{n}(2\pi)^{n/2}}\exp\left\{\frac{-1}{2\sigma^{2}}\left[A+n(\bar{y}-\mu)^2\right]\right\}\quad,\quad where\ A=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

The joint Posterior distribution of μ & σ is given by

$$P(\mu,\sigma\ \,|\,y)\propto P(\sigma).L(\mu,\sigma\,|\,Y)$$

$$\Rightarrow P(\mu,\sigma\ \,|\,y)\propto\frac{1}{\sigma^{c}}.\frac{1}{\sigma^{n}(2\pi)^{n/2}}\exp\left\{\frac{-1}{2\sigma^{2}}\left[A+n(\bar{y}-\mu)^2\right]\right\}$$

$$\Rightarrow P(\mu,\sigma\ \,|\,y)=k\,\frac{1}{\sigma^{n+c}}\exp\left\{\frac{-1}{2\sigma^{2}}\left[A+n(\bar{y}-\mu)^2\right]\right\}$$

(5.3.1)

$$Where\qquad K^{-1}=\int_{0}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sigma^{n+c}}\exp\left\{\frac{-1}{2\sigma^{2}}\left[A+n(\bar{y}-\mu)^2\right]\right\}d\mu\,d\sigma$$

$$=\int_{0}^{\infty}\frac{\exp\left(-A/2\sigma^{2}\right)}{\sigma^{n+c}}\left\{\int_{--\infty}^{\infty}\exp\left[\frac{-n(\bar{y}-\mu)^2}{2\sigma^{2}}\right]d\mu\right\}d\sigma$$

$$=\int_{0}^{\infty}\frac{\exp\left(-A/2\sigma^{2}\right)}{\sigma^{n+c}}\left\{\frac{\sqrt{2\pi}\sigma}{\sqrt{n}}\right\}d\sigma$$

$$=\frac{\sqrt{2\pi}}{\sqrt{n}}\int_{0}^{\infty}\frac{\exp\left(-A/2\sigma^{2}\right)}{(2\sigma)^{\frac{1}{2}(n+c-1)}}(2)^{\frac{1}{2}(n+c-1)}d\sigma$$

$$=\frac{\sqrt{2\pi}}{\sqrt{n}}\int_{0}^{\infty}\frac{\exp\left(-A/2\sigma^{2}\right)}{\sigma^{n+c-1}}d\sigma$$

Put $2\sigma^2 = k \;\Rightarrow\; \sigma = \frac{\sqrt{k}}{\sqrt{2}} \Rightarrow 4\sigma\, d\sigma = dk$ or $d\sigma = \frac{\sqrt{2}\, dk}{4\sqrt{k}}$

$$\Rightarrow \qquad K^{-1} = \frac{\sqrt{2\pi}}{\sqrt{n}} \int_0^\infty \frac{\exp(-A/k)\; 2^{\frac{1}{2}(n+c-1)}}{k^{\frac{1}{2}(n+c-1)}} \frac{\sqrt{2}}{4\sqrt{k}}\, dk$$

$$= \sqrt{\frac{2\pi}{n}}\; 2^{\frac{1}{2}(n+c)-2} \int_0^\infty \frac{\exp(-A/k)}{k^{\frac{1}{2}(n+c-1)+1}}\, dk$$

$$= \sqrt{\frac{2\pi}{n}}\; 2^{\frac{1}{2}(n+c-4)} \frac{\Gamma(n+c-2)/2}{A^{\frac{n+c-2}{2}}}$$

$$\Rightarrow \qquad K = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}}$$

$$\therefore P(\mu, \sigma \mid y) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{1}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{y}-\mu)^2\right]\right\}$$

(5.3.2)

The marginal posterior of σ is given by integrating out μ in (5.3.2) we have

$$P(\sigma \mid y) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c}} \int_{-\infty}^{\infty} \exp\left\{\frac{-n}{2\sigma^2}(\bar{y}-\mu)^2\right\} d\mu$$

$$= \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c}} \frac{\sqrt{2\pi}\;\sigma}{\sqrt{n}}$$

$$= \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}}$$

(5.3.3)

Bayes estimator of σ is given by

$$\hat{\sigma} = E(\sigma \mid y) = \int_0^\infty \sigma\; P(\sigma \mid y)\, d\sigma$$

$$=\int_0^\infty \sigma \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}} \, d\sigma$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-2}} \, d\sigma$$

$$=\frac{A^{\frac{n+c-2}{2}} \; 2^{\frac{1}{2}(n+c-2)}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{(2\sigma^2)^{(n+c-2)/2}} \, d\sigma$$

substituting $2\sigma^2 = z$, we have

$$\hat{\sigma} = \frac{A^{\frac{n+c-2}{2}} \; 2^{\frac{1}{2}(n+c-2)-\frac{3}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/z\right\}}{(z)^{\frac{(n+c-3)}{2}+1}} \, dz$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}}} \frac{\Gamma(n+c-3/2)}{A^{\frac{n+c-3}{2}}}$$

$$=\sqrt{\frac{A}{2}} \frac{\Gamma\left(\frac{n+c-3}{2}\right)}{\Gamma\left(\frac{n+c-2}{2}\right)}$$

(5.3.4)

Bayes estimator of σ^2 is

$$\hat{\sigma}^2 = E\left(\sigma^2 \mid y\right) = \int_0^\infty \sigma^2 \; P(\sigma \mid y) \, d\sigma$$

$$=\int_0^\infty \sigma^2 \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}} \, d\sigma$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-3}} \, d\sigma$$

substituting $2\sigma^2 = t$, we have

$$\hat{\sigma}^2 = \frac{A^{\frac{n+c-2}{2}}\; 2^{\frac{1}{2}(n+c-3)}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\{-A/t\}}{(t)^{\frac{(n+c-4)}{2}+1}} \frac{\sqrt{2}}{4\sqrt{t}}\, dt$$

$$= \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2} \frac{\Gamma(n+c-4/2)}{A^{\frac{n+c-4}{2}}}$$

$$= \frac{A}{n+c-4}$$

(5.3.5)

If we put c=4 in (4.5) we observe that MLE of σ^2 coincides with $\hat{\sigma}^2$ and for c=3, the UMVUE of σ^2 is the same as Bayes estimate for σ^2.

Now, the marginal distribution of μ is given by

$$P(\mu \mid y) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left[\frac{-1}{2\sigma^2}\left\{A + n(\bar{y}-\mu)^2\right\}\right]}{\sigma^{n+c}}\, d\sigma$$

Put $2\sigma^2 = r$, we have

$$P(\mu \mid y) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}} \sqrt{2}}{\Gamma\left(\frac{n+c-2}{2}\right)} \int_0^\infty \frac{\exp\left[\frac{-1}{r}\left\{A + n(\bar{y}-\mu)^2\right\}\right]}{(r)^{\frac{n+c-1}{2}+1}}\, dr$$

$$= \sqrt{\frac{n}{\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)} \frac{\Gamma\left(\frac{n+c-1}{2}\right)}{\left\{A + n(\bar{y}-\mu)^2\right\}^{\frac{n+c-1}{2}}}$$

$$= \sqrt{\frac{n}{\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)} \frac{\Gamma\left(\frac{n+c-1}{2}\right)}{A^{\frac{n+c-1}{2}}\left\{1 + \frac{n(\bar{y}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}$$

$$=\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\left\{1+\frac{n(\bar{y}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}\ ;\ -\infty<\mu<\infty$$

(5.3.6)

Bayes estimator of μ is given by

$$\hat{\mu}=E(\mu\,|\,y)=\int_{-\infty}^{\infty}\mu\,P(\mu\,|\,y)d\mu$$

$$\hat{\mu}=\int_{-\infty}^{\infty}\mu\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\left\{1+\frac{n(\bar{y}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}d\mu$$

$$\hat{\mu}=\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)}\int_{-\infty}^{\infty}\frac{\mu}{\left\{1+\frac{n(\bar{y}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}d\mu$$

substituting $\frac{n(\bar{y}-\mu)^2}{A}=\frac{t^2}{n+c-2}$ we have

$$\hat{\mu}=\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)}\int_{-\infty}^{\infty}\frac{\bar{y}-\frac{\sqrt{A/n}\ t}{\sqrt{n+c-2}}}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}}\left(\frac{-\sqrt{A/n}}{\sqrt{n+c-2}}\right)dt$$

$$\hat{\mu}=\frac{-1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\sqrt{n+c-2}}\int_{-\infty}^{\infty}\frac{\bar{y}-\frac{\sqrt{A/n}\ t}{\sqrt{n+c-2}}}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}}\,dt$$

$$=\frac{-\bar{y}}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\sqrt{n+c-2}}\int_{-\infty}^{\infty}\frac{1}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}}dt+\frac{\sqrt{A/n}}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)}\int_{-\infty}^{\infty}\frac{t}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}}dt$$

Since $\frac{\sqrt{A/n}}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)}\int_{-\infty}^{\infty}\frac{t}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}}dt$ is an odd function equal to zero.

$$\therefore \hat{\mu} = \frac{-\bar{y}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{1}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} dt$$

$$= \frac{-\bar{y}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{-\sqrt{n+c-2}\sqrt{A/n}}{\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}} d\mu = \bar{y}$$

(5.3.7)

5.4 Bayesian intervals for parameter of normal distribution:

The joint posterior of μ and σ^2 is given by

$$P(\mu, \sigma^2 \mid y) = \frac{k}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{y}-\mu)^2\right]\right\}; \quad -\infty < \mu < \infty; \sigma > 0$$

(5.4.1)

where k is normalizing constant.

Putting c=2 in the (5.4.1) we have

$$P(\mu, \sigma^2 \mid y) = \frac{k}{\sigma^{n+2}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{y}-\mu)^2\right]\right\}$$

where $A = \sum_{i=1}^{n}(y_i - \bar{y})^2$ and $k^{-1} = \sqrt{\frac{\pi}{n}}\left(\frac{2}{A}\right)^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)$

(5.4.2)

Integrating out μ and restoring the normalizing constant k, the marginal posterior density for σ^2 is given by

$$P(\sigma^2 \mid y) = \left(\frac{A}{2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\exp\left(-A/2\sigma^2\right)}{\left(\sigma^2\right)^{(n-1/2)+1}}$$

(5.4.3)

Similarly we obtain the marginal posterior of μ

$$P(\mu \mid y) = \frac{\sqrt{n}}{\sqrt{A}\,\beta\left(\frac{1}{2}, \frac{n-1}{2}\right)} \frac{1}{\left[1+\frac{n(\bar{y}\_\mu)^2}{A}\right]^{\frac{n}{2}}}$$

(5.4.4)

from (5.4.3) it follows A/σ^2 is distributed as χ^2 with (n-1) degrees of freedom. For $1-\alpha$ equal tail credible interval $[T_1, T_2]; T_1 < T_2$ must satisfy the conditions

$$\frac{\alpha}{2} = \int_0^{T_1} \Phi(\sigma^2 \mid y)\, d\sigma^2 = \int_{T_2}^{\infty} \Phi(\sigma^2 \mid y)\, d\sigma^2$$

$$\frac{\alpha}{2} = P(T_2 \le \sigma^2)$$

$$1-\frac{\alpha}{2} = P\left(\chi^2 \le \frac{A}{T_2} = \chi^2_{1-\frac{\alpha}{2}}\right)$$

Hence $\dfrac{A}{T_2} = \chi^2\left(1-\dfrac{\alpha}{2}, n-1\right)$ or $T_2 = \dfrac{A}{\chi^2\left(1-\dfrac{\alpha}{2}, n-1\right)}$

where $\chi^2(k, v)$ = upper 100% point of a χ^2 distribution with v degrees of freedom.

Similarly

$$\frac{\alpha}{2} = P(\sigma^2 \le T_1) = P\left(\chi^2 \ge \frac{A}{T_1}\right)$$

which implies $\dfrac{A}{T_1} = \chi^2\left(\dfrac{\alpha}{2}, n-1\right)$ or $T_1 = \dfrac{A}{\chi^2\left(\dfrac{\alpha}{2}, n-1\right)}$

Thus the $(1-\alpha)$ equal tail credible interval of σ^2 is

$$\left[\frac{A}{\chi^2\left(\frac{\alpha}{2}, n-1\right)}, \frac{A}{\chi^2\left(1-\frac{\alpha}{2}, n-1\right)}\right]$$

(5.4.5)

which is the same as the classical $(1-\alpha)$ confidence interval. The posterior distribution of σ^2 in (5.4.3) is unimodal. Hence the shortest credible interval & the HPD interval are the same.

The $(1-\alpha)$-HPD interval $[H_1, H_2]$ must simultaneously satisfy

$$P\left(\frac{A}{H_2} < \chi^2 < \frac{A}{H_1}\right) = 1-\alpha$$

(5.4.6)

& $$\exp\left[\frac{-A}{2}\left(\frac{1}{H_1}-\frac{1}{H_2}\right)\right]=\left(\frac{H_1}{H_2}\right)^{\frac{n+1}{2}}$$

(5.4.7)

It follows from (5.4.4) that $\frac{\sqrt{n-1}\,(\mu-\bar{y})}{\sqrt{A/n}}$ follows Student's t-distribution with (n-1) degrees of freedom. Also the posterior distribution of μ is unimodal and symmetric about $\bar{y}$. Hence the $(1-\alpha)$ equal-tail credible, shortest credible & the HDP intervals for μ are identical. Such an interval $[H_1, H_2]$ must satisfy the condition.

$$1-\alpha=P(H_1<\mu<H_2)=P\left[\frac{\sqrt{n-1}(H_1-\bar{y})}{\sqrt{A/n}}<t<\frac{\sqrt{n-1}(H_2-\bar{y})}{\sqrt{A/n}}\right]$$

$$=P\left[-t\left(\frac{\alpha}{2},n-1\right)<t<t\left(\frac{\alpha}{2},n-1\right)\right]$$

$$H_1=\bar{y}-\frac{s}{\sqrt{n}}t\left(\frac{\alpha}{2},n-1\right);\ H_2=\bar{y}+\frac{s}{\sqrt{n}}t\left(\frac{\alpha}{2},n-1\right)$$

(5.5.8)

where $t(k,m)$=100% point of student's t- distribution with m degrees of freedom. Here again we observe that the $(1-\alpha)$-HDP intervals of μ is the same as the classical $(1-\alpha)$ confidence interval for μ.

Example: we generated a random sample of size n=20 from a normal dist with μ =20, σ^2 =3. For this sample $\bar{x}$ =20.50, A=1733.25.

We want to construct the 90% credible & HPD intervals for μ & σ^2.

The 90% equal tail credible limits for T_1, T_2 are given by (5.4.5):

$$T_1=\frac{A}{\chi^2\left(\frac{\alpha}{2},n-1\right)}=\frac{173.25}{\chi^2(0.05,19)}=\frac{173.25}{30.14}=5.75$$

$$T_2=\frac{A}{\chi^2\left(1-\frac{\alpha}{2},n-1\right)}=\frac{173.25}{\chi^2(0.95,19)}=\frac{173.25}{10.12}=17.12$$

The width of the interval I_c =11.37. The 90% of the HPD (as well as the shortest credible) limits H_1 and H_2 which satisfies (5.5.6);(5.5.7) are given by

$H_1 = 4.77$, $H_2 = 14.38$. The width of the interval $I_H = 9.61 < I_C$ as anticipated. The 90% HPD (as well as the equal- tail & the shortest credible) limits for μ are given by (5.5.8)

$$\bar{y} \pm t\left(\frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n}} = 20.50 \pm t(5\%,19)\frac{s}{\sqrt{20}}$$

$$= 20.50 \pm 1.17 = 19.33\ ,\ 21.67$$

90% equal tail credible interval for $\sigma^2 = [5.75\ , 17.12]$

90% HPD & shortest credible for $\sigma^2 = [4.77, 14.38]$

90% HPD & shortest credible & equal-tail credible interval for $\mu = [1933,\ 21.67]$

5.5 Normal Approximation for normal distribution:

The pdf of normal distribution is given by

$$f\left(y \mid \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \qquad ; -\infty < y < \infty\ ; -\infty < \mu < \infty$$

The likelihood function is given by

$$L\left(\mu, \sigma^2 \mid y\right) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}} \exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}\right]$$

Consider the prior $P\left(\mu, \sigma^2\right) = 1$

Therefore posterior density is given by

$$P\left(\mu, \sigma^2 \mid y\right) \propto p\left(\mu, \sigma^2\right) L\left(\mu, \sigma^2 \mid y\right)$$

$$P\left(\mu, \sigma^2 \mid y\right) \propto \frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}} \exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}\right]$$

$$\log P\left(\mu, \sigma^2 \mid y\right) = \log constant - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}$$

(5.5.1)

differentating $= n$ (5.5.1) *partially w.r.t* μ, σ^2 *we have*

$$\frac{\partial \log P\left(\mu, \sigma^2 \mid y\right)}{\partial \mu} = \frac{\sum_{i=1}^{n}(y_i - \mu)}{\sigma^2}$$

$$\Rightarrow \frac{\partial \log P(\mu,\sigma^2 \mid y)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^4}$$

Posterior mode is $\hat{\mu} = \bar{y}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2 = \frac{(n-1)s^2}{n}$

$$\text{or } \frac{-\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial \mu^T \, \partial \mu} = \frac{n}{\sigma^2}$$

$$\Rightarrow \frac{\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial \sigma^{2^T} \, \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(\bar{y}-\mu)^2 = \frac{n}{2\sigma^4} - \frac{n\sigma^2}{2\sigma^4} = \frac{-n}{2\sigma^4}$$

$$\therefore \frac{-\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial \sigma^{2^T} \, \partial \sigma^2} = \frac{n}{2\sigma^4}$$

$$\therefore I(\mu,\sigma^2) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$

$$\therefore I^{-1}(\mu,\sigma^2) = \frac{1}{n^2/2\sigma^6}\begin{pmatrix} n/2\sigma^4 & 0 \\ 0 & n/\sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}$$

$$\therefore I^{-1}(\hat{\mu},\hat{\sigma}^2) = \begin{pmatrix} \hat{\sigma}^2/n & 0 \\ 0 & 2\hat{\sigma}^4/n \end{pmatrix}$$

Hence $P(\mu,\sigma^2 \mid y) \sim N\left(\hat{\theta}, \begin{pmatrix} \hat{\sigma}^2/n & 0 \\ 0 & 2\hat{\sigma}^4/n \end{pmatrix}\right)$

where $\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \frac{(n-1)s^2}{n} \end{pmatrix}$

Now consider $P(\mu,\sigma^2) = \frac{1}{\sigma^2}$

The likelihood is given by

$$L(\mu,\sigma^2 \mid y) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}} \exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}\right]$$

Therefore the posterior density of $P(\mu,\sigma^2 \mid y)$ is given by

$$P(\mu,\sigma^2 \mid y) \propto p(\mu,\sigma^2)\, L(\mu,\sigma^2 \mid y)$$

$$P(\mu,\sigma^2 \mid y) \propto \frac{1}{\sigma^2}\frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}}\exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}\right]$$

$$\therefore \log P(\mu,\sigma^2 \mid y) = \log constant - \left(\frac{n}{2}+1\right)\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

(5.5.1)

$$\frac{\partial \log P(\mu,\sigma^2 \mid y)}{\partial\mu} = \frac{\sum_{i=1}^{n}(y_i-\mu)}{\sigma^2}$$

$$\frac{\partial \log P(\mu,\sigma^2 \mid y)}{\partial\sigma^2} = -\left(\frac{n}{2}+1\right)\frac{1}{\sigma^2} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^4}$$

Posterior mode is $\hat{\mu} = \bar{y}$ *and* $\hat{\sigma}^2 = \frac{1}{n+2}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{(n-1)s^2}{n+2}$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$

$$\frac{\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial\mu^T\,\partial\mu} = \frac{-n}{\sigma^2}$$

$$\text{or } \frac{-\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial\mu^T\,\partial\mu} = \frac{n}{\sigma^2}$$

$$\frac{\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial\sigma^{2T}\,\partial\sigma^2} = \left(\frac{n}{2}+1\right)\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i-\mu)^2 = \frac{n+2}{2\sigma^4} - \frac{n\sigma^2}{\sigma^4} = \frac{-n+2}{2\sigma^4}$$

$$\Rightarrow \frac{-\partial^2 \log P(\mu,\sigma^2 \mid y)}{\partial\sigma^{2T}\,\partial\sigma^2} = \frac{n-2}{2\sigma^4}$$

$$\therefore I(\mu,\sigma^2) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n-2/2\sigma^4 \end{pmatrix}$$

$$\Rightarrow I^{-1}(\mu,\sigma^2) = \frac{1}{\frac{n}{\sigma^2}\cdot\frac{n-1}{2\sigma^4}}\begin{pmatrix} n/2\sigma^4 & 0 \\ 0 & n/\sigma^2 \end{pmatrix} = \frac{2\sigma^6}{n(n-2)}\begin{pmatrix} n-2/2\sigma^4 & 0 \\ 0 & n/\sigma^2 \end{pmatrix}$$

$$\Rightarrow I^{-1}(\mu,\sigma^2) = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^2/n-2 \end{pmatrix}$$

$$\therefore I^{-1}\left(\hat{\mu},\hat{\sigma}^2\right)=\begin{pmatrix}\hat{\sigma}^2/n & 0\\ 0 & 2\hat{\sigma}^4/n-2\end{pmatrix}$$

$$\therefore P\left(\mu,\sigma^2 \mid y\right)\sim N\left(\hat{\theta}, I^{-1}\left(\hat{\theta}\right)\right)$$

where $\hat{\theta}=\begin{pmatrix}\hat{\mu}\\ \hat{\sigma}^2\end{pmatrix}=\begin{pmatrix}\bar{y}\\ \dfrac{(n-1)s^2}{n+2}\end{pmatrix}$ and $I^{-1}\left(\hat{\theta}\right)=\begin{pmatrix}\hat{\sigma}^2/n & 0\\ 0 & 2\hat{\sigma}^4/n-2\end{pmatrix}$

where $\hat{\sigma}^2=\dfrac{(n-1)s^2}{n+2}$

5.6 Selection of Prior Distribution for Normal Distribution:

Let us consider the normal distribution with known mean μ & unknown variance σ^2. Bernardo (2005) gave an objective Bayesian decision theoretic solution to point estimation of the normal variance with mean as unknown & behavior of solution found is compared from both a Bayesian & a frequentists perspective. Sinha (1998) has obtained 95% predictive intervals for various sets of hyper parameters using sample size n=100 from Mendenhall & Harder (1958) mixture model. Lee (1997) derived a suitable conjugate prior (universe chi-squared distribution) for the normal variance with mean as known quantity. Evans (1964) derived some general forms of estimators of the variance of normal distribution. Using Bayesian methods & the conditions under which they lead to previously proposed Geodman (1960) estimators.

We use the following informative priors for find the posterior_distribution for the unknown parameter variance σ^2 and also find the posterior predictive distributions under these informative priors which are given below:

1) Inverse chi-square distribution (conjugate prior).
2) Inverse gamma distribution (conjugate prior).
3) Levy distribution.
4) Gumbel type=II distribution.

Let $y_1, y_2, ..., y_n$ be a random sample from the normal distribution with parameters mean μ (known) and variance σ^2 (unknown).

The likelihood function of the sample observations $Y: y_1, y_2, ..., y_n$ is

$$L(Y \mid \sigma^2)=\prod_{i=1}^{n} f\left(y_i \middle| \mu,\sigma\right)=\prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-1}{2\sigma^2}\left(y_i-\mu\right)^2\right\}\right]$$

$$\Rightarrow L(Y \mid \sigma^2) = \left(\frac{1}{\sigma^2 2\pi}\right)^{n/2} \exp\left\{\frac{-w}{2\sigma^2}\right\}$$

(5.6.1)

$$where\ w = \sum_{i=1}^{n} (y_i - \mu)^2$$

5.7 The Posterior Distribution of σ^2 Using Inverse Chi-Squared Distribution as prior:

It is assumed that the prior distribution of σ^2 is an inverse chi-squared distribution with hyper parameters $'a_1'$and $'b_1'$ which is given below:

$$P_1(\sigma^2) = \frac{b_1^{\frac{a_1}{2}}}{2^{\frac{a_1}{2}} \Gamma\left(\frac{a_1}{2}\right)} (\sigma^2)^{\frac{-a_1}{2}-1} \exp\left\{\frac{-b_1}{2\sigma^2}\right\}, \sigma^2 > 0; a_1, b_1 > 0$$

(5.7.1)

The density kernel is

$$P_1(\sigma^2) \propto (\sigma^2)^{\frac{-a_1}{2}-1} \exp\left\{\frac{-b_1}{2\sigma^2}\right\}$$

(5.7.2)

Now the posterior distribution of the parameter σ^2 for the given data $Y : y_1, y_2, \ldots.., y_n$ is

$$P_1(\sigma^2 \mid Y) \propto L(Y; \sigma^2) p_1(\theta)$$

$$\Rightarrow P_1(\sigma^2 \mid Y) \propto (\sigma^2)^{\frac{-n}{2}} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{\frac{-a_1}{2}-1} e^{\frac{-b_1}{2\sigma^2}}$$

$$\Rightarrow P_1(\sigma^2 \mid Y) \propto (\sigma^2)^{\frac{-\alpha_1}{2}-1} e^{\frac{-\beta_1}{2\sigma^2}}$$

(5.7.3)

which is the density kernel of the inverse chi-squared distribution with parameters: $\alpha_1 = a_1 + n$ and $\beta_1 = b_1 + w$. So the posterior distribution of parameter σ^2 for the given data is an inverse chi-squared distribution having parameters α_1 and β_1 where α_1 and β_1 have already been defined above.

5.8 The Posterior Distribution of σ^2 Using Inverted Gamma Distribution as Prior:

Now the prior distribution of σ^2 is assumed to be the inverted gamma distribution with the hyper parameters $'a_2'$ and $'b_2'$ having the following pdf

$$P_2(\sigma^2) = \frac{b_2^{a_2}}{\Gamma(a_2)} (\sigma^2)^{-(a_2+1)} \exp\left\{\frac{-b_2}{\sigma^2}\right\}, \sigma^2 > 0; a_2, b_2 > 0$$

(5.8.1)

Now the posterior distribution of the parameter σ^2 for given data $Y: y_1, y_2, \ldots\ldots, y_n$ is

$$P_2(\sigma^2 \mid Y) \propto (\sigma^2)^{-(a_2+1)} e^{\frac{-b_2}{\sigma^2}} (\sigma^2)^{\frac{-n}{2}} e^{\frac{-w}{2\sigma^2}}$$

$$\Rightarrow P_2(\sigma^2 \mid Y) \propto (\sigma^2)^{-\left(a_2+\frac{n}{2}+1\right)} e^{\frac{-1}{\sigma^2}\left(\frac{2b_2+w}{2}\right)}$$

$$\Rightarrow P_2(\sigma^2 \mid Y) \propto (\sigma^2)^{-(\alpha_2+1)} e^{\frac{-\beta_2}{\sigma^2}}$$

(5.8.2)

which is the density kernel of the inverted gamma distribution with the parameters $\alpha_2 = a_2 + \frac{n}{2}$ and $\beta_2 = \frac{(2b_2 + w)}{2}$. so the posterior distribution of parameter σ^2 for the given data is an inverted gamma (α_2, β_2) where α_2 and β_2 has been defined above.

5.9 The Posterior Distribution of σ^2 Using Levy Distribution as Prior:

Third prior distribution is assumed to be Levy distribution with hyper parameter $'b_3'$ which has the following pdf

$$P_3(\sigma^2) = \sqrt{\frac{b_3}{2\pi}} (\sigma^2)^{\frac{-3}{2}} e^{\frac{-b_3}{2\sigma^2}} \; ; \sigma^2 > 0, b_3 > 0$$

(5.9.1)

Now the posterior distribution of the parameter σ^2 for given data $Y: y_1, y_2, \ldots\ldots, y_n$ is

$$P_3(\sigma^2 \mid Y) \propto (\sigma^2)^{-\left(\frac{n}{2}\right)} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{-\left(\frac{3}{2}\right)} e^{\frac{-b_3}{2\sigma^2}}$$

$$\Rightarrow P_3(\sigma^2 \mid Y) \propto (\sigma^2)^{-(\alpha_3+1)} e^{\frac{-\beta_3}{\sigma^2}} \qquad (5.9.2)$$

which is the density kernel of the inverted gamma distribution with the parameters $\alpha_3 = \frac{n+1}{2}$ and $\beta_3 = \frac{(b_3 + w)}{2}$. so the posterior distribution of parameter σ^2 for the given data is an inverted gamma (α_3, β_3) where α_3 and β_3 has been already defined above.

5.10 The Posterior Distribution of σ^2 Using Gumbel Type-II Distribution as Prior:

The Gumbel Type-II distribution with the hyper parameters $'a_4'$and$'b_4'$ is supposed to be the fourth prior distribution of σ^2 which is:

$$P_4(\sigma^2) = a_4 b_4 (\sigma^2)^{-(a_4+1)} e^{-b_4(\sigma^2)^{-a_4}} \quad ; \quad \text{where } \sigma^2 > 0,\ a_4, b_4 > 0$$

For making the conjugate prior, we take $a_4 = 1$ then the prior is:

$$P_4(\sigma^2) = b_4 (\sigma^2)^{-2} e^{-b_4(\sigma^2)^{-1}}$$

(5.10.1)

Now the posterior distribution of the parameter σ^2 for given data $Y: y_1, y_2, \ldots\ldots, y_n$ is

$$P_4(\sigma^2 \mid Y) \propto (\sigma^2)^{-\left(\frac{n}{2}\right)} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{-2} e^{\frac{-b_4}{\sigma^2}}$$

$$\Rightarrow P_4(\sigma^2 \mid Y) \propto (\sigma^2)^{-(\alpha_4+1)} e^{\frac{-\beta_4}{\sigma^2}}$$

(5.10.2)

which is the density kernel of the inverted gamma distribution with the parameters $\alpha_4 = \frac{n}{2} + 1$ and $\beta_4 = \frac{(2b_4 + w)}{2}$. so the posterior distribution of parameter σ^2 for the given data is an inverted gamma (α_4, β_4) where α_4 and β_4 has been already defined above.

5.11 The Posterior Predictive Distribution:

We observe that there are two types of posterior distributions which are derived under all priors. So we now derive posterior predictive distributions under these posterior distributions i.e. inverted gamma and inverse chi-squared distributions.

a) The Posterior Predictive Distribution under the Prior Inverse Chi-squared Distribution:

The posterior predictive distribution for $X = Y_{n+1}$ given that $Y: y_1, y_2, \ldots\ldots, y_n$

under posterior inverse chi-squared distribution is:

$$P_1(X \mid Y) = \int_0^{\infty} p(x \mid \sigma^2)\, P_1(\sigma^2 \mid Y)\, d\sigma^2$$

$$\Rightarrow P_1(X \mid Y) = \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \frac{\beta_1^{\alpha_1/2} (\sigma^2)^{\frac{-\alpha_1}{2}-1}}{2^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1}{2}\right)} \exp\left\{\frac{-\beta_1}{2\sigma^2}\right\} d\sigma^2$$

$$P_1(X \mid Y) = \frac{\beta_1^{\alpha_1/2}}{2^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1}{2}\right) (2\pi)^{1/2}} \int_0^{\infty} \frac{1}{(\sigma^2)^{\frac{\alpha_1}{2}+\frac{1}{2}+1}} \exp\left\{\frac{-1}{\sigma^2}\left(\frac{(x-\mu)^2}{2} + \frac{\beta_1}{2}\right)\right\} d\sigma^2$$

$$P_1(X \mid Y) = \frac{\beta_1^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1+1}{2}\right)}{2^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1}{2}\right) (2\pi)^{1/2} \left[\frac{\beta_1}{2} + \frac{(x-\mu)^2}{2}\right]^{\frac{\alpha_1+1}{2}}}$$

$$P_1(X \mid Y) = \frac{\beta_1^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1+1}{2}\right) 2^{\frac{\alpha_1+1}{2}} \left[1 + \frac{(x-\mu)^2}{\alpha_1(\beta_1/\alpha_1)}\right]^{\frac{-(\alpha_1+1)}{2}}}{2^{\alpha_1/2}\, \Gamma\left(\frac{\alpha_1}{2}\right) (2\pi)^{1/2}\, \beta_1^{\frac{\alpha_1+1}{2}}}$$

$$\Rightarrow P_1(X \mid Y) = \frac{\Gamma\left(\frac{\alpha_1+1}{2}\right) \left[1 + \frac{(x-\mu)^2}{\alpha_1(\beta_1/\alpha_1)}\right]^{\frac{-(\alpha_1+1)}{2}}}{\sqrt{\frac{\beta_1}{\alpha_1}}\, \sqrt{\alpha_1 \pi}\, \Gamma\left(\frac{\alpha_1}{2}\right)}$$

(5.11.1)

which is the probability density function of t-distribution i.e.

$$(X \mid Y) \sim t\left(\alpha_1, \mu, \frac{\beta_1}{\alpha_1}\right);\ -\infty < x, \mu < \infty\ ; \alpha_1 > 0$$

(5.11.2)

Hence $(X \mid Y)$ has the t-distribution with three parameters $u_1, v_1,$ and w_1

Where $u_1 = \alpha_1$, $v_1 = \mu$ and $w_1 = \frac{\beta_1}{\alpha_1}$; $w_1 > 0$

b) The Posterior Predictive Distribution under the Prior Inverted Gamma Distribution:

The posterior predictive distribution for $X = Y_{n+1}$ given that $Y: y_1, y_2, \ldots.., y_n$

under posterior inverted gamma distribution is:

$$P_2(X\mid Y)=\frac{\Gamma\left(\alpha_2+\frac{1}{2}\right)\left[1+\frac{(y-\mu)^2}{2\alpha_1(\beta_2/\alpha_2)}\right]^{\frac{-(2\alpha_2+1)}{2}}}{\sqrt{(\beta_2/\alpha_2)}\sqrt{2\alpha_2\pi}\,\Gamma\left(\frac{2\alpha_2}{2}\right)}$$

(5.11.3)

which is the probability density function of t-distribution i.e.

$$X\mid Y\sim t\left(2\alpha_2,\mu,\frac{\beta_2}{\alpha_2}\right);-\infty<y,\mu<\infty\,;\alpha_2>0$$

(5.11.4)

Hence $X|Y$ has the t-distribution with three parameters u_2, v_2, and w_2

where $u_2=2\alpha_2,\ v_2=\mu$ *and* $w_2=\frac{\beta_2}{\alpha_2}$; $w_2>0$.

c) The Posterior Predictive Distribution under the Prior Levy Distribution:

The posterior predictive distribution for $X=Y_{n+1}$ given that $Y:y_1,y_2,.....,y_n$ under posterior inverted gamma distribution is:

$$P_3(X\mid Y)=\int_0^{\infty}p(x\mid\sigma^2)\;p_3(\sigma^2\mid Y)d\sigma^2$$

$$\Rightarrow P_3(X\mid Y)=\frac{\Gamma\left(\alpha_3+\frac{1}{2}\right)\left[1+\frac{(x-\mu)^2}{2\alpha_3(\beta_3/\alpha_3)}\right]^{\frac{-(2\alpha_3+1)}{2}}}{\sqrt{(\beta_3/\alpha_3)}\sqrt{2\alpha_3\pi}\,\Gamma\left(\frac{2\alpha_3}{2}\right)}$$

(5.11.5)

which is the probability density function of t-distribution i.e.

$$X\mid Y\sim t\left(2\alpha_3,\mu,\frac{\beta_3}{\alpha_3}\right);-\infty<x,\mu<\infty\,;\alpha_3>0$$

(5.11.6)

Hence $X|Y$ has the t-distribution with three parameters u_3, v_3, and w_3

where $u_3=2\alpha_3,\ v_3=\mu$ *and* $w_3=\frac{\beta_3}{\alpha_3}$; $w_3>0$.

d) The Posterior Predictive Distribution under the Prior Gumbel Type-II Distribution:

The posterior predictive distribution for $X = Y_{n+1}$ given that $Y : y_1, y_2, ..., y_n$ under posterior inverted gamma distribution is:

$$P_4(X \mid Y) = \frac{\Gamma\left(\alpha_4 + \frac{1}{2}\right)\left[1 + \frac{(x-\mu)^2}{2\alpha_4(\beta_4/\alpha_4)}\right]^{\frac{-(2\alpha_4+1)}{2}}}{\sqrt{(\beta_4/\alpha_4)}\sqrt{2\alpha_4\pi}\,\Gamma\left(\frac{2\alpha_4}{2}\right)}$$

(5.11.7)

$$X \mid Y \sim t\left(2\alpha_4, \mu, \frac{\beta_4}{\alpha_4}\right); \qquad -\infty < x, \mu < \infty \,; \alpha_4 > 0$$

Hence $X \mid Y$ has the t-distribution with three parameters $u_4, v_4, \textit{and } w_4$

where $u_4 = 2\alpha_4,\ v_4 = \mu \textit{ and } w_4 = \frac{\beta_4}{\alpha_4}$; $w_4 > 0$

5.12 Comparison of priors with respect to posterior variances:

The variances of the posterior distributions are calculated and are given in Table 5.2.1, 5.2.2, 5.2.3:

1. For the posterior inverse chi-square distribution we have

$$V(\sigma^2 \mid Y) = \frac{2\beta_1^{\ 2}}{(\alpha_1 - 2)^2(\alpha_1 - 4)} \text{ ;provided } \alpha_1 > 4$$

2. For the posterior inverted gamma distribution we have

$$V(\sigma^2 \mid Y) = \frac{\beta_i^{\ 2}}{(\alpha_i - 1)^2(\alpha_i - 2)} \qquad \text{provided } \alpha_i > 2, i =,2,3,4$$

5.13 Comparison using the posterior predictive variances:

The posterior predictive variances using different prior distributions are given in the tables 5.2.4, 5.2.5 and 5.2.6.

The posterior predictive variances under inverse chi-square as prior distribution is

$$\mathrm{V}(\sigma^2 \mid \mathrm{Y}) = \frac{\beta_1}{\alpha_1 - 2}; \alpha_1 > 2$$

and the posterior predictive variances under the inverted gamma, Levy and Gumbel typeII distributions as priors is

$$V(\sigma^2 \mid Y) = \frac{2\beta_i}{\alpha_i - 2}; \alpha_i > 2 : i = 2,3,4$$

Example 5.1(Gelman et al, 1995): Simon Newcomb set up an experiment in 1882 to measure the speed of light. Newcomb measured the amount of time required for light to travel a distance of 7442 meters. The measurements are given below:

28, 26, 33, 24, 34, -44, 27, 16, 40, -2, 29, 22, 24, 21, 25, 30, 23, 29, 31, 19, 24, 20, 36, 32, 36, 28, 25, 21, 28, 29, 37, 25, 28, 26, 30, 32, 36, 26, 30, 22, 36, 23, 27, 27, 28, 27, 31, 27, 26, 33, 26, 32, 32, 24, 39, 28, 24, 25, 32, 25, 29, 27, 28, 29, 16, 23.

We apply the normal model, assuming that all 66 measurements are independent draws from a normal distribution with mean μ and variance σ^2. We use the following programme for obtaining the posterior mode and posterior standard error for μ and σ under different priors and are shown in table 5.1.1.

#Bayesian analysis of normal distribution with different priors in R.

#Prior=1.

```
pos.normal<-function(theta,y)

{

z<-(y-theta[1])/theta[2]

n<-length(y)

lik<- n*log(theta[2])+sum(z^2)

pri<--log(1)

pos<-pri+lik

return(pos)

}

speed<-
c(28,29,24,37,36,26,29,26,22,20,25,23,32,27,33,24,36,28,2
7,32,28,24,21,32,26,27,24,29,34,25,36,30,28,39,16,44,30,2
8,32,27,2823,27,23,25,36,31,24,16,29,21,26,27,25,40,31,28
,30,26,32,     -2,19,29,22,33,25)

out<-nlm(pos.normal,y=speed,c(15,12),hessian=T)

std.err<-sqrt(diag(solve(out$hessian)))
```

```
> out
$minimum
[1] 212.0851
$estimate
[1] 26.21211 15.08062
$hessian
              [,1]          [,2]
[1,]  5.804110e-01 -9.987634e-05
[2,] -9.987634e-05  5.801214e-01
> std.err
[1] 1.312599 1.312927
```

#Prior=1/sigma.

```
pos.normal<-function(theta,y)
{
z<-(y-theta[1])/theta[2]
n<-length(y)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/theta[2])
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,y=speed,c(15,12),hessian=T)
std.err<-sqrt(diag(solve(out$hessian)))
> out
$minimum
[1] 214.7947
$estimate
```

```
[1] 26.21211 14.96764
$hessian
              [,1]          [,2]
[1,]  0.5892058182 -0.0001023543
[2,] -0.0001023543  0.5978357346
> std.err
[1] 1.302766 1.293329
```

#Prior=1/sigma^2.

```
pos.normal<-function(theta,y)
{
z<-(y-theta[1])/theta[2]
n<-length(y)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^2))
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,y=speed,c(15,12),hessian=T)
std.err<-sqrt(diag(solve(out$hessian)))
> out
$minimum
[1] 217.4969
$estimate
[1] 26.21211 14.85718
$hessian
              [,1]          [,2]
[1,]  0.5979996113 -0.0001043341
```

```
[2,] -0.0001043341  0.6158139449
> std.err
[1] 1.293152 1.274310
```

 # Prior=1/sigma^3.

```
pos.normal<-function(theta,y)
{
z<-(y-theta[1])/theta[2]
n<-length(y)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^3))
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,y=speed,c(15,12),hessian=T)
std.err<-sqrt(diag(solve(out$hessian)))
> out
$minimum
[1] 220.1917
$estimate
[1] 26.21211 14.74913
$hessian
              [,1]          [,2]
[1,]  0.6067937002 -0.0001066644
[2,] -0.0001066644  0.6340591994
> std.err
[1] 1.283747 1.255842
```

 # Prior=1/sigma^4.

```
pos.normal<-function(theta,y)
{
z<-(y-theta[1])/theta[2]
n<-length(y)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^4))
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,y=speed,c(15,12),hessian=T)
std.err<-sqrt(diag(solve(out$hessian)))
> out
$minimum
[1] 222.8793
$estimate
[1] 26.21212 14.64341
$hessian
             [,1]         [,2]
[1,]  0.615586710 -0.000110063
[2,] -0.000110063  0.652567967
> std.err
[1] 1.274546 1.237904
```

Table 5.1.1: Posterior mode and Posterior standard error of parameters of Normal distribution with different priors.

Prior	Posterior mode Mu	Posterior Std.err Mu	Posterior mode sigma	Posterior Std.err sigma
1	26.21211	1.312599	15.08062	1.312927
1/sigma	26.21211	1.302766	14.96764	1.293329
1/(sigma^2)	26.21211	1.293152	14.85718	1.274310
1/(sigma^3)	26.21211	1.283747	14.74913	1.255842
1/(sigma^4)	26.21211	1.274546	14.64341	1.237904

Example: 5.2 (simulation): We generated a sample of size 30, 60, 100 from normal pdf with parameter μ and σ^2 to represent small, moderate and large sample sizes. Also we have taken different values for parameters and hyper parameters.

Programme for simulation in R-software:

 # Simulations in R Software for posterior variance

 # Posterior variance of sigma^2 under chi-square as a prior

```
sim.var <-function(a1,b1,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha1<-(a1+n);beta1<-b1+w

pvc<-2*(beta1^2)/(((alpha1-2)^2)*(alpha1-4))

return(pvc)

 }

a1=b1=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))

y3<-rnorm(100,20,sqrt(6))

cbind(sim.var(a1,b1,mu,y1),sim.var(a1,b1,mu,y2),sim.var(a
1,b1,mu,y3))
```

Posterior variance of sigma^2 under inverted gamma as a prior

```
sim.var <-function(a2,b2,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha2<-(a2+n/2);beta2<-(2*b2+w)/2

pvg<-(beta2^2)/(((alpha2-1)^2)*(alpha2-2))
```

```
return(pvg)

 }

a2=b2=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))

y3<-rnorm(100,20,sqrt(6))

cbind(sim.var(a2,b2,mu,y1),sim.var(a2,b2,mu,y2),sim.var(a
2,b2,mu,y3))
```

 # Posterior variance of sigma^2 under levy distribution as a prior

```
sim.var <-function(a3,b3,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha3<-(1+n/2);beta3<-(b3+w)/2

pvl<-(beta3^2)/(((alpha3-1)^2)*(alpha3-2))

return(pvl)

 }

a3=b3=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))

y3<-rnorm(100,20,sqrt(6))

cbind(sim.var(a3,b3,mu,y1),sim.var(a3,b3,mu,y2),sim.var(a
3,b3,mu,y3))
```

 # Posterior variance of sigma^2 under Gumbel type II distribution as a prior

```
sim.var <-function(a4,b4,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha4<-(1+n/2);beta4<-(2*b4+w)/2

pvgb<-(beta4^2)/(((alpha4-1)^2)*(alpha4-2))

return(pvgb)

 }

a4=b4=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))
```

```
y3<-rnorm(100,20,sqrt(6))

cbind(sim.var(a4,b4,mu,y1),sim.var(a4,b4,mu,y2),sim.var(a4,b4,mu,y3))
```

Table5.2.1: Variances of the posterior distribution of σ^2 using different priors with n=30,60&100 mean=20, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters a_i=b_i=c_i	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	V1	5	0.18889	0.14535	0.25304	0.29984
		10	0.14535	0.09637	0.29984	0.40534
		15	0.11654	0.07045	0.35060	0.52671
		20	0.09637	0.05482	0.40534	0.66396
		25	0.08162	0.04453	0.46404	0.81708
		30	0.07045	0.03731	0.52671	0.98607
		35	0.06175	0.03200	0.59335	1.17094
		40	0.05482	0.02795	0.66396	1.37167
		45	0.04918	0.02478	0.73853	1.58829
		50	0.04918	0.02222	0.81708	1.82077
60	V2	5	0.94822	0.66542	1.27028	1.37267
		10	0.66542	0.37790	1.37267	1.58936
		15	0.49165	0.24371	1.47903	1.82192
		20	0.37790	0.17095	1.58936	2.07035
		25	0.29968	0.12723	1.70365	2.33466
		30	0.24371	0.09894	1.82192	2.61484
		35	0.20235	0.07956	1.94415	2.91089
		40	0.17095	0.06568	2.07035	3.22281
		45	0.14656	0.05539	2.20052	3.55061
		50	0.14656	0.04753	2.33466	3.89428
100	V3	5	0.40182	0.29359	0.53830	0.60564
		10	0.29359	0.17886	0.60564	0.75224
		15	0.22503	0.12236	0.67696	0.91471
		20	0.17886	0.09025	0.75224	1.09305
		25	0.14626	0.07015	0.83149	1.28727
		30	0.12236	0.05665	0.91471	1.49736
		35	0.10428	0.04710	1.00190	1.72332
		40	0.09025	0.04005	1.09305	1.96516
		45	0.07913	0.03468	1.18818	2.22287
		50	0.07913	0.03047	1.28727	2.49645

Table 5.2.2: Variances of the posterior distribution of σ^2 using different priors with n=30,60&100 mean=25, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters $a_i=b_i=c_i$	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	V1	5	0.2523937	0.1097682	0.2946667	0.4012673
		10	0.1899811	0.07588661	0.3450102	0.4596842
		15	0.1494729	0.05722005	0.3993219	0.5220694
		20	0.121592	0.04559845	0.4576019	0.5884228
		25	0.10151	0.03774458	0.5198501	0.6587445
		30	0.086512	0.03211649	0.5860666	0.7330344
		35	0.0749788	0.02790248	0.6562513	0.8112926
		40	0.0658896	0.02463815	0.7304043	0.8935191
		45	0.058579	0.02204008	0.8085255	0.9797138
		50	0.0525954	0.01992622	0.890615	1.069877
60	V2	5	0.5914351	0.2691619	0.9145326	0.6001036
		10	0.4868972	0.1956175	0.9443692	0.6490119
		15	0.4072712	0.1488878	0.9746847	0.699836
		20	0.3453757	0.1174243	1.005479	0.7525757
		25	0.2964068	0.09525345	1.036752	0.8072311
		30	0.257062	0.07904755	1.068505	0.8638023
		35	0.2250157	0.06683971	1.100736	0.9222892
		40	0.1985958	0.05740935	1.133446	0.9826918
		45	0.1765768	0.04996766	1.166635	1.04501
		50	0.1580456	0.043987	1.200303	1.109244
100	V3	5	0.7697074	0.7045608	0.5548802	0.7478029
		10	0.6774931	0.555792	0.5655727	0.7727143
		15	0.6001549	0.4479326	0.5763672	0.7980339
		20	0.5347553	0.3676098	0.5872638	0.8237616
		25	0.4790326	0.3064124	0.5982624	0.8498975
		30	0.4312254	0.2588601	0.609363	0.8764415
		35	0.3899469	0.2212718	0.6205657	0.9033937
		40	0.3540953	0.1911091	0.6318705	0.9307541
		45	0.322787	0.1665803	0.6432772	0.9585227
		50	0.2953072	0.1463948	0.654786	0.9866994

Table 5.2.3: Variances of the posterior distribution of σ^2 using different priors with n=30,60&100 mean=30, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters $a_i=b_i=c_i$	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	V1	5	0.2092032	0.1791297	0.5126912	0.4455196
		10	0.1596833	0.115501	0.5784639	0.5723825
		15	0.1271546	0.0826537	0.6482048	0.7151184
		20	0.1045247	0.06324058	0.721914	0.8737273
		25	0.08806729	0.05067085	0.7995914	1.048209
		30	0.07566947	0.04198114	0.8812371	1.238564
		35	0.06605869	0.03567106	0.9668511	1.444792
		40	0.05842996	0.03091072	1.056433	1.666893
		45	0.05225295	0.0272086	1.149984	1.904867
		50	0.04716612	0.02425725	1.247502	2.158714
60	V2	5	0.3433044	0.3144186	0.3779776	0.606197
		10	0.2858744	0.2269046	0.3972446	0.6553481
		15	0.2417584	0.1715936	0.4169906	0.706415
		20	0.2071842	0.1345361	0.4372154	0.7593976
		25	0.1796133	0.1085437	0.4579192	0.8142958
		30	0.1572906	0.08962644	0.479102	0.8711098
		35	0.1389733	0.07543399	0.5007636	0.9298395
		40	0.1237629	0.06451246	0.5229042	0.9904849
		45	0.1109973	0.05592516	0.5455237	1.053046
		50	0.1001806	0.04904738	0.5686221	1.117523
100	V3	5	0.9624093	0.5631733	0.775842	0.7773726
		10	0.8457255	0.4458316	0.7884762	0.8027677
		15	0.7479776	0.3605443	0.8012124	0.828571
		20	0.6654127	0.2968758	0.8140506	0.8547825
		25	0.5951436	0.2482518	0.8269909	0.8814021
		30	0.534923	0.2103816	0.8400332	0.9084299
		35	0.4829835	0.1803789	0.8531775	0.9358659
		40	0.4379213	0.15625	0.8664239	0.96371
		45	0.3986116	0.1365856	0.8797724	0.9919623
		50	0.3641455	0.120369	0.8932228	1.020623

The results obtained using above programme are presented in tables 5.2.1 :5.2.2; 5.2.3 for different values of hyper parameters, n and mean. In the above Tables 5.2.1, 5.2.2 and 5.2.3, it is observed that the values of the posterior predictive variances under inverted gamma distribution using different values of hyper parameters are less as compare to other priors which means we can prefer the prior inverted gamma distribution as a prior for the variance of normal distribution.

 # Simulations in R Software for predictive distribution

 # Predictive Posterior variance of sigma^2 under chi-square as a prior

```
pre.var <-function(a1,b1,mu,y){

n<-length(y)

w<-sum((y-mu)^2)

alpha1<-(a1+n)

beta1<-b1+w

pvc<-beta1/(alpha1-2)

return(pvc)

 }

a1=b1=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4));y3<-rnorm(100,20,sqrt(6))

cbind(pre.var(a1,b1,mu,y1),pre.var(a1,b1,mu,y2),pre.var(a1,b1,mu,y3))
```

 # Predictive Posterior variance of sigma^2 under inverted gamma as a prior

```
pre.var <-function(a2,b2,mu,y){

n<-length(y)

w<-sum((y-mu)^2)

alpha2<-a2+n/2

beta2<-(2*b2+w)/2

pvg<-beta2/(alpha2-1)

return(pvg)

 }

a2=b2=5

mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4));y3<-rnorm(100,20,sqrt(6))
```

```
cbind(pre.var(a1,b1,mu,y1),pre.var(a1,b1,mu,y2),pre.var(a
1,b1,mu,y3))
```

 # Predictive Posterior variance of sigma^2 under levy distribution as a prior

```
pre.var <-function(a3,b3,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha3<-(1+n/2);beta3<-(b3+w)/2

pvl<-(beta3)/(alpha3-1)

return(pvl)

 }

a3=b3=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))

y3<-rnorm(100,20,sqrt(6))

cbind(pre.var(a3,b3,mu,y1),pre.var(a3,b3,mu,y2),pre.var(a
3,b3,mu,y3))
```

 # Predictive Posterior variance of sigma^2 under Gumbel type II distribution as a prior

```
pre.var <-function(a4,b4,mu,y){

n<-length(y); w<-sum((y-mu)^2)

alpha4<-(1+n/2);beta4<-(2*b4+w)/2

pvgb<-(beta4)/(alpha4-1)

return(pvgb)

 }

a4=b4=5
mu<-20
y1<-rnorm(30,20,sqrt(2));y2<-rnorm(60,20,sqrt(4))

y3<-rnorm(100,20,sqrt(6))

cbind(pre.var(a4,b4,mu,y1),pre.var(a4,b4,mu,y2),pre.var(a
4,b4,mu,y3)
```

5.2.4: Variances of the posterior predictive distribution of σ^2 using different priors with n=30,60&100 mean=20, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters $a_i=b_i=c_i$	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	V1	5	1.849804	1.583599	2.587101	2.541914
		10	1.737988	1.462016	2.753768	2.875247
		15	1.652175	1.382358	2.920435	3.20858
		20	1.58424	1.326129	3.087101	3.541914
		25	1.529123	1.284318	3.253768	3.875247
		30	1.483509	1.252009	3.420435	4.20858
		35	1.445135	1.226294	3.587101	4.541914
		40	1.412405	1.20534	3.753768	4.875247
		45	1.384158	1.187939	3.920435	5.20858
		50	1.359532	1.173256	4.087101	5.541914
60	V2	5	3.884797	2.994491	3.69685	3.484049
		10	3.672679	2.738787	3.780183	3.650716
		15	3.489619	2.541198	3.863516	3.817383
		20	3.330028	2.383933	3.94685	3.984049
		25	3.189665	2.255791	4.030183	4.150716
		30	3.065252	2.149368	4.113516	4.317383
		35	2.954217	2.059573	4.19685	4.484049
		40	2.854512	1.982793	4.280183	4.650716
		45	2.764487	1.916388	4.363516	4.817383
		50	2.682798	1.858389	4.44685	4.984049
100	V3	5	5.095442	4.361218	5.162707	5.377846
		10	4.905839	4.076369	5.212707	5.477846
		15	4.733014	3.836028	5.262707	5.577846
		20	4.574835	3.630519	5.312707	5.677846
		25	4.429517	3.452781	5.362707	5.777846
		30	4.295551	3.297542	5.412707	5.877846
		35	4.171658	3.160783	5.462707	5.977846
		40	4.056743	3.039391	5.512707	6.077846
		45	3.949864	2.930913	5.562707	6.177846
		50	3.850207	2.833392	5.612707	6.277846

Table 5.2.5: Variances of the posterior predictive distribution of σ^2 using different priors with n=30,60&100 mean=25, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters $a_i=b_i=c_i$	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	V1	5	2.650895	1.960673	1.96791	2.257561
		10	2.433672	1.760533	2.301244	2.424228
		15	2.266966	1.629407	2.634577	2.590894
		20	2.13499	1.536847	2.96791	2.757561
		25	2.027916	1.46802	3.301244	2.924228
		30	1.939302	1.414836	3.634577	3.090894
		35	1.864755	1.372506	3.96791	3.257561
		40	1.80117	1.338015	4.301244	3.424228
		45	1.746295	1.309369	4.634577	3.590894
		50	1.698456	1.2852	4.96791	3.757561
60	V2	5	3.053136	2.941562	3.436798	3.638667
		10	2.90217	2.692644	3.603464	3.72200
		15	2.771884	2.500298	3.770131	3.805334
		20	2.658302	2.347206	3.936798	3.888667
		25	2.558404	2.222465	4.103464	3.97200
		30	2.469859	2.118866	4.270131	4.055334
		35	2.390834	2.031455	4.436798	4.138667
		40	2.319873	1.956712	4.603464	4.22200
		45	2.255802	1.892069	4.770131	4.305334
		50	2.197663	1.835609	4.936798	4.388667
100	V3	5	4.488128	4.329209	4.77395	4.838322
		10	4.326641	4.047072	4.87395	4.888322
		15	4.179445	3.80902	4.97395	4.938322
		20	4.044722	3.605468	5.07395	4.988322
		25	3.920953	3.429423	5.17395	5.038322
		30	3.806853	3.275662	5.27395	5.088322
		35	3.701333	3.140206	5.37395	5.138322
		40	3.603458	3.019969	5.47395	5.188322
		45	3.512428	2.912524	5.57395	5.238322
		50	3.427549	2.815932	5.67395	5.288322

Table 5.2.6: Variances of the posterior predictive distribution of σ^2 using different priors with n=30,60&100 mean=30, variances V1=2, V2=4 & V3=6.

Size	σ^2	Hyper Parameters $a_i=b_i=c_i$	Inverse Chi-Square Prior	Inverted Gamma Prior	Levy Prior	Gumbel Type-II Prior
30	**V1**	5	2.628931	1.816164	1.894188	1.967029
		10	2.414598	1.64613	2.060854	2.300362
		15	2.250109	1.456092	2.227521	2.633695
		20	2.119890	1.397619	2.394188	2.967029
		25	2.014240	1.352435	2.560854	3.300362
		30	1.926805	1.316472	2.727521	3.633695
		35	1.853249	1.287169	2.894188	3.967029
		40	1.790510	1.262833	3.060854	4.300362
		45	1.736366	1.242299	3.227521	4.633695
		50	1.689163	1.223460	3.394188	4.967029
60	**V2**	5	3.336274	3.154604	3.688514	3.696962
		10	3.164489	2.878373	3.771848	3.863629
		15	3.016237	2.664922	3.855181	4.030295
		20	2.886991	2.495032	3.938514	4.196962
		25	2.773317	2.356603	4.021848	4.363629
		30	2.67256	2.241636	4.105181	4.530295
		35	2.582637	2.144634	4.188514	4.696962
		40	2.501891	2.061689	4.271848	4.863629
		45	2.428983	1.989953	4.355181	5.030295
		50	2.362827	1.927298	4.438514	5.196962
100	**V3**	5	4.991427	4.64504	5.04143	5.190872
		10	4.806638	4.336139	5.09143	5.290872
		15	4.638203	4.075503	5.14143	5.390872
		20	4.484042	3.85264	5.19143	5.490872
		25	4.342414	3.659894	5.24143	5.590872
		30	4.211851	3.491546	5.29143	5.690872
		35	4.091105	3.34324	5.34143	5.790872
		40	3.979108	3.211597	5.39143	5.890872
		45	3.874944	3.093959	5.44143	5.990872
		50	3.777817	2.988204	5.49143	6.090872

The results obtained using above programme are presented in tables 5.2.4 :5.2.5; 5.2.6 for different values of hyper parameters, n and mean. In the above Tables 5.2.4, 5.2.5 and 5.2.6, it is observed that the values of the posterior predictive variances under inverted gamma distribution using different values of hyper parameters are less as compare to other priors which means we can prefer the prior inverted gamma distribution as a prior for the variance of normal distribution.

BIBLIOGRAPHY

1. Abramowitz, M. And Stegun, I.(Ed). (1964), Handbook Of Mathematical Functions, National Bureau Of Standards, Applied Mathematical Series 55,U.S.Government Printing Office.
2. Barlett, M. S.(1922): Properties Of Sufficiency & Statistical Tests. Proc. R. Soc. London A, 160,268-282.
3. Bazovsky, I. (1961): Reliability Theory And Practice, Prentice Hall, New Jersey.
4. Berger, J. O. (1985): Statistical Decision Theory And Bayesian Analysis. 2^{nd} Edn.
5. Bernardo, J. M. (2005): Intrinsic Point Estimation Of The Normal Variance. New Dehli: Anamay Pub.
6. Bernardo,J.M And Smith, A. F. M(1994): Bayesian Theory , J.Wiley And Sons, New York
7. Bernoulli, J.(1713): Ars Conjectandi,Basilea: Thurnisius.[11.19]
8. Box, G.E.P., And Tiao, G.C. (1973): Bayesian Inference In Statistical Analysis. Addison Wesley.
9. Cox,D.R., and Reid,N.,(1987): Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society,B,49,1-39.
10. Davis, D. J. (1952): “The Analysis Of Some Failure Data”, J.Amer. Statist. Assoc. 47, 113 150.
11. De Finitte (1970): Teoria Delle Probabilitia, Turin: Einandi, English Translation As Theory Of Probability In 1975, Chichester: Wiley.
12. De Groot, M.H.(1970): Optimal Statistical Decisions, Mc Graw Hills, New York.
13. Diciccio, T.J.(1986): Approximate Conditional Inference For Location Families, Candian Journal Statistics 14,5-18
14. Edwards,W.,Lindman, H. Savage, L.J.(1963): Bayesian Statistical Inference For Psychological Research, Psychological Review, 70, 193-242.
15. Erans,I.G.(1976): Unpublished Lecture Notes.University College Of

Wales,.Aberyswyth, U.K.

16. Erdelyi,A., (1953) Higher Transcendental Functions, vol2 McGraw Hill, New York.

17. Evans, I. G. (1964). Bayesian Estimation Of The Variance Of A Normal Distribution. Journal Of The Royal Statistical Society. Series B (Methodological), 26, 63-85.

18. Feldman,D., And Fox,H. (1968): Estimation Of The Parameter N In The Binomial Distribution. Journal Of The American Statistical Association, 63,150-158.

19. Fisher, R. A.(1922): On The Mathematical Foundations Of Theoretical Statistics, Phil.Trans,Roy. Soc. Series A. Vol.222.Pp-308-358.

20. Gauss, C. F. (1809): Theoria Motus Corporum Coelestium, Hamsburd: Perths & Besser.(English Translation By C.H. Davis, Published 1857, Boston : Little Brown, Co.).

21. Gauss, C. F. (1816):Bestimmuing Der Genauigkeit Der Beobachtugen, Zeit-Schrift Astronomi, I, 185-1.

22. Gelman, A.,Carlin, J.B.,Stern, H.S.,And B.,.R.D.(1995): Bayesian Data Analysis Chapman B Hall,London.

23. Geodman, L. A. (1960). A Note On The Estimation Of Variance", Sankhyz, 22, 221-228.

24. Gianola,D., and Fernando,R.L.,(1986): Bayesian Methods in Animal Breeding Theory.J.Annual.Sci.63,217-244.

25. Haldane, J. B. S.(1931): A Note On Inverse Probability Proceedings Of Cambridge Philosophical Society 28,55- 61.

26. Hartigan, J. A.(1964): Invariant Prior Distribution Annals Of Mathematical Statistics, 35, 836- 845.

27. Hoff, P. D (2009): A First Course in Bayesian Statistical Methods. Springer

28. Irwin, J.O., 1959: On The Estimation Of The Mean Of A Poisson Istribution From A Sample With The Zero Class Missing .Biometrics, 15,324-326.

29. Jefferys, H. (1946): An Invariant Form For The Prior Probability In Estimation Problems; Proceedings Of The Royal Society Of London, Series.A,186,453-461

30. Jefferys, R.C.(1983): The Logic Of Decision, II Edition. Univesity Of Chicago Press.

31. Jeffrey, H.(1939): Theory Of Probability . Oxford University Press.

32. Johnson, R.A. (1970): Asymptotic Expansions Associated With Posterior Distributions. Annals Of Mathematical Statistics, 41, 196-215

33. Laplace, P.S. (1774): Memoire Sur La Probabilities Des Causes Parles Evenemens. Men .Acad. R. Sci.,6, 621-656. Translated By Stephen M.Stigler And Reprinted Translation In Statistical Science, 1986,Z(3, 359-378).

34. Laplace,P.S, (1812): Theorie Analytiquie Des Probabilities, Paris: Courair.

35. Lee, P. M. (1997). Bayesian Statistics An Introduction. John Wiley & Sons Inc. New York. Toronto, 50-53,278-279.

36. Legendre, A.(1805): Nouelles Methods Pour La Determination Des Orbites Dis Cometes, Coucier, Pairs.

37. Lindley, D.V. (1965): Introduction To Probability And Statistics Fom A Bayesian View Point(Parti An II) Cambridge University Press.

38. Mendenhall,W. And Hader, R.J(1958): Estimation Of Parameters Of Mixed Exponentially Distributed Failure Time Distribution From Censored Failure Data, Biometrika 45,504-520.

39. Morvie, A. De.(1738): Approximatio Ad Summan Ferminorum Binomii $(a+b)^n$ In Seriem Expansi, Supplementum Ii To Miscellanae Analytica, 1-7.

40. Mosteller, F. And Wallace, D.L. (1964): Applied Bayesian And Classical Inference. Springer-Verlag, New York

41. Pratt, J.W. (1965). Bayesian interpretation of standard inference statements (with discussion). Journal of the Royal Statistical Society, B 27, 169-203

42. Poisson, S.D.(1837): Recherches Sur La Probabilitie Des Jugements En Matiere Criminelle Et En Matiere Civile, Precedecs Des Regles Generals Du Calcul Des Probabilities.410.Pp.I X + 415. Paris.
43. Raffia. H And Schlaifer, R. (1961): Applied Statistical Decision Theory. Division Of Research, Graduate School Of Business Administration , Harvard University.
44. Royall, R. M.(1997): Statistical Evidence- A Likelihood Paradigm, Chapman & Hall.
45. Sinha, S.K. (1998). Bayesian Estimation . New Age International (P) Limited, Publishers,
46. Stigler,S.M.(1987):Who Discovered Bayes Theorem? American Statistics 37:290-296.
47. Tierney, L., Kass, R.E.,And Kadane,J.B.(1989a): Approximate Marginal Densities Of Non –Linear Functions", Biometrika,76,425-433.
48. Tierney, L., Kass, R.E.,And Kadane,J.B.(1989a): Approximate Marginal Densities Of Non –Linear Functions", Biometrika,76,425-433.
49. Yoichi Miyata (2004): "Fully Exponential Laplace Approximations Using Asymptotic Modes", Journal Of The American Statistical Association, 99,1037-1049.
50. Zellner, A.(1977): Maximal Data Information Prior Distribution: In: New Developments In The Application Of Bayesian Methods , A.Aykac & C, Brumat (Eds.), North Holland Publishing Co.,201-15.
51. Zellner, A.(1971):An Introduction to Bayesian Inference in Econometrics.Wiley, NewYork.
52. Edwards, W., H. Lindman, and L.J. Savage, (1963). Bayesian statistical inference for psychological research. Psychological Review, 70, 193-242.
53. Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society, B, 20, 102-107.
54. Berger, J.O. and R. Wolpert, (1984). The Likelihood Principle. Harvard, California:Institute of Mathematical Statistics

55. Rubin, D.B., and N. Schenker, (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. Sociological Methodologies, 16, 131-144.

56. Clogg, C.C., D.B. Rubin, N. Schenker, B. Schultz, and L. Wideman, (1991). Multiple imputation of industry and occupation codes in Census public-use samples using

57. Bayesian logistic regression. Journal of the American Statistical Association, 86, 68-78.

58. Freedman, L.S., D.J. Spiegelhalter, and M.K.B. Parmar, (1994). The what, why and how of Bayesian clinical trials monitoring. Statistics in Medicine, 13, 1371-1383.

59. Khan, A.A. (1997). Asymptotic Bayesian Analysis in Location-Scale Models. Ph.D. Thesis submitted to the Dept. of Mathematics and Statistics, HAU, Hisar.