CONSTRUCTION OF POSTERIOR DENSITIES USING ASYMPTOTIC ASSUMPTIONS

Dissertation Submitted

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

By
HUMMARA SULTAN

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
Tel.:0194-2416589(Res.)
Telefax:0194-2421357

Dated: _____________

CERTIFICATE

This is to certify that the scholar **Ms. Hummara Sultan,** has carried out the present dissertation entitled **"Construction of Posterior Densities Using Asymptotic Assumptions"** 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

*In the completion of this dissertation, there are countless debits incurred that can never be adequately repaid, yet the most appropriate thing is to acknowledge. First of all my thanks goes to my esteemed supervisor **Dr. Sheikh Parvaiz Ahmad** whose keen interest, clear thinking, sympathetic nature and insistence on better quality inspired me to work hard.*

*I wish to express deep sense of gratitude to **Prof.Aquil. Ahmad** (Head, Department of statistics). It is all because of his open mindedness, objective way of thinking, deep insight in research that I came to analyze, interpret in a proper perspective. Without his sincere devotion, constant encouragement, valuable guidance and sustained interest in my work, the study would not have been completed.*

*I wish to place my affectionate thanks to **Dr. M.A.K Baig, Dr. Tariq Rashid Jan** for their help and valuable suggestions.*

No words would be sufficient to express my gratitude to all my fellow scholars Raja Sultan Ahmad Reshi, Adil Rashid, Adil Hamid Khan, Nusrat Mushtaq, Javeed Ahmad, Shabir Ahmad Chopan, for their constant support, encouragement and memorable company.

A very special thanks to my sister Sheikh Sumaira, my brother Sheikh Mohammad Shoaib for their love and support and to my friends and colleagues Liyaqat Ali, Sabrina Ashraf, Saniya Jalil, Sheikh Bilquees. This research work would not have been possible without their contribution, as they extended their valuable assistance in my field work.

Last but not the least, my very special thanks to my dearest parents for all their efforts and challenges which they have faced for bringing me where I stand today and for bearing all the inconveniences during the completion of present work.

Hummara Sultan

PREFACE

Bayesian approach to statistical inference exploits the idea that the only satisfactory description of uncertainty is by means of probability. Bayesian statistics is an approach in which estimates are based on a synthesis of a prior distribution and current sample data. When significant prior is available, the Bayesian approach shows how to utilize it sensibly. 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 four chapters:

Chapter I includes introduction to Bayesian statistics, Bayes theorem, sequential nature of Bayes theorem, likelihood to Bayesian analysis, marginal and conditional inferences, prior and some important types of priors. Normal and Laplace methods of approximation for posterior modes and some important models like Exponential, Two Parameter Exponential, Gamma and Normal distributions are also discussed.

Chapter II is devoted to the Bayesian estimation for exponential distribution under different priors. Laplace and Normal approximations to the posterior density of exponential distribution are also discussed. We have also discussed Bayesian estimation for two parameter exponential distribution. To illustrate the methods, we have developed some programs in S-PLUS for numerical and graphical representation of posterior densities.

Chapter III deals with the estimation of parameters of Gamma distribution with complete sample. Normal and Laplace approximation to the posterior density of Gamma distribution are also discussed. The methods are illustrated with the help of some programs developed in S-PLUS for numerical and graphical representation of posterior densities.

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

DEDICATED

TO MY

BELOVED

GRAND FATHER

LATE MOHAMMAD SULTAN KHAN

CONTENTS

CHAPTER – 1

INTRODUCTION

1.1 Introduction:

Uncertainty plays an important role in our lives. A satisfactory description of uncertainty is by means of probability. The probability is a powerful tool of maintaining, understanding, and controlling this important feature of our appreciation of our environment. Bayesian approach to statistical inference exploits the idea that the only satisfactory description of uncertainty is by means of probability. Bayesian statistics is an approach in which estimates are based on a synthesis of a prior distribution and current sample data. Bayesian statistics requires the mathematics of probability 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 de Laplace (1812), Jefferys (1939) and de Finetti (1970) are actually entitled "probability theory". Indeed, Bayesian methods (i) reduce statistical inference to problems in probability theory, thereby minimizing the need for completely new concepts, and (ii) serve to discriminate among conventional statistical techniques either providing a logical justification to some (and making explicit the conditions which they are valid) or proving the logical in consistency of others.

Bayesian statistics have been used to deal with a wide variety of problems in many scientific and engineering areas. Whenever a quantity is to be inferred, or some conclusion is to be drawn, from observed data, Bayesian principles and tools can be used. The idea that forms the basis of the Bayesian approach is as:

i) Since we are uncertain about the true value of the parameters, we will consider them to be random variables.
ii) The rules of probability are used directly to make inferences about the parameters.
iii) Probability statements about parameters must be interpreted as "degree of belief". The prior distribution must be subjective.
iv) We revise our beliefs about parameters after getting the data by using Bayes theorem. This gives our posterior distribution which gives the relative weights to each parameter value after analyzing the data.

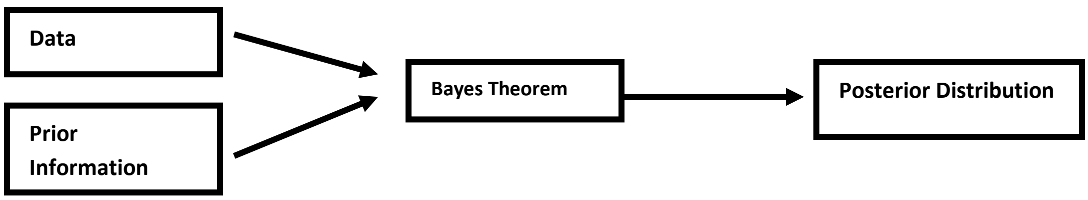
Bayesian statistics is predictive, unlike conventional frequentist statistics. This means we can easily find the conditional probability distribution of the next observation given the sample data. Bayesian approach to statistics is very different from the classical methodology, it formally seeks use of prior information and Bayes theorem provides the basis for making use of this information. When significant prior 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 possible. 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 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 persons 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 procedures and is likely to be more reasonable for moderate and especially small sample sizes where non Bayesian procedures break down (e.g., Berger 1985).

1.2 Bayes theorem:

Bayesian analysis is based upon a theorem first developed by an 18th century English mathematician, logician, and clergy man Thomas Bayes (1701-1761). He developed the theorem in his study of the theory of logic and inductive reasoning. The theorem provides a mathematical basis for relating the degree to which an observation (or new information) confirms the various hypothesized causes or state of nature. His major mathematical works, including the theorem, were published in 1763. Later, in 1774 the theorem was proved independently by Laplace. Bayes theorem is an essential element of the Bayesian approach to statistical inference is the direct qualification 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



Suppose that $X' = x_1, x_2, ..., x_n$ is a vector of n observations whose probability distribution $P(X \mid \theta)$ depends upon the values of k parameters $\theta' = \theta_1, \theta_2, ..., \theta_k$. Suppose also that θ itself has a probability distribution $P(\theta)$. Then,

$$P(X \mid \theta)P(\theta) = P(X, \theta) = P(\theta \mid X)P(X).$$

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

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)} \qquad (1.2.1)$$

Also we can write

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

$$= \sum P(X \mid \theta)P(\theta); \qquad \theta \, discrete$$

Where the sum or the integral is taken over the admissible range of θ, and where E indicates averaging with respect to distribution of θ (e.g., Box and Tiao, 1973;

Gelman, Carlin, Stern and Rubin,1995; Lee,1997 and Carlin and Louis,2000). Thus we may write (1.2.1) alternatively as

$$P(\theta \mid X) \propto P(X \mid \theta)P(\theta) \tag{1.2.2}$$

which is 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 $P(X \mid \theta)$ is likelihood function of θ which represents the contribution of X(data) to knowledge about θ (e.g., Berger,1985 and Zellner, 1971). Correspondingly, $P(\theta \mid X)$, which tells us what is known about θ given knowledge of the data X, is called the posterior distribution of θ given X. The quantity 'k' is a normalizing constant.

The term 'Bayesian' however, came into use only around 1950 and in fact it is not clear that Bayes' would endorsed the very broad interpretation of probability now called "Bayesian". Laplace independently proved a more general version of Bayes' theorem and put it to good use in solving problems in celestial mechanics, medical statistics and, by some accounts, even jurisprudence.

1.3 Sequential Nature of Bayes' Theorem:

Now given the data X, $P(X \mid \theta)$ in (1.2.2) may be regarded as a function not of X but of θ. When so regarded, following Fisher (1922), it is called the likelihood function of θ for given X and can be written as $L(\theta \mid X)$.We can thus write Bayes formula as

$$P(\theta \mid X) = L(\theta \mid X)P(\theta)$$

(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 X_1, then Bayes initial formula gives,

$$P(\theta \mid X_1) \propto P(\theta)L(\theta \mid X_1)$$

(1.3.2)

Now suppose we have a second sample of observation X_2, distributed independently of first sample, then

$$P(\theta \mid X_1, X_2) \propto P(\theta)L(\theta \mid X_1)L(\theta \mid X_2)$$

$$\propto P(\Theta \mid X_1)L(\Theta \mid X_2) \qquad (1.3.3)$$

The expression (1.3.3) precisely of the same form as (1.3.2) except that $P(\theta \mid X_1)$, the posterior distribution for θ given X_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 mth stage the likelihood associated with the mth observation is combined with the posterior distribution of θ after m-1 observations to give the new posterior distribution.

$$P(\theta \mid X_1, X_2, ..., X_m) \propto P(\theta \mid X_1, X_2, ..., X_{m-1})L(\theta \mid X_m) : m = 1,2,......., n \qquad (1.3.4)$$

where $P(\Theta \mid X_1) \propto P(\Theta)L(\Theta \mid X_1)$.

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 an Tiao,1973).

1.4 From Likelihood to Bayesian Analysis:

An informal summary of the likelihood principle may be that inferences from data to hypothesis should depend on how likely the actual data are under competing hypothesis, not on how likely imaginary data would have been under a single “null” hypothesis or any other properties of merely possible data.

A more precise interpretation may be that inference procedures which make inferences about simple hypothesis should not be justified by appealing to probabilities assigned to observations that have not occurred. The usual interpretation is that any two probability models with the same likelihood function yield the same inference for θ. Some authors mistakenly claim that frequentist inference, such as the use of maximum likelihood estimation (MLE), obeys the likelihood, though it does not. Some argue that, although the subject of priors gets more attention, the true contention between frequentist and Bayesian inference is the likelihood principle, which Bayesian inference obeys, and frequentist inference does not. Some Bayesians have argued that Bayesian inference is incompatible with the likelihood principle on the grounds that there is no such thing as an isolated likelihood function (Bayarri and DeGroot, 1987). They argue that in a Bayesian analysis there is no principled distinction between the likelihood function and the prior probability function.

Although the likelihood principle is implicit in Bayesian statistics, it was developed as a separate principle by Barnard (Barnard 1949), and became a focus of interest when Birnbaum (1962) showed that it followed from the widely accepted sufficiency and conditionality principles (Bernardo and Smith 2000). Using Bayes' rule with a chosen probability model means that the data X affect posterior inference only through the function $L(X\mid\theta)$, which, when regarded as a function of θ, for fixed X, is called the `likelihood function'. In this way Bayesian inference obeys what is sometimes called the `likelihood principle', which states that for a given sample of data, any two probability models $L(X\mid\theta)$ that have the same likelihood function yield the same inference for θ (Bernardo and Smith, 2000 and Gelman et.al. 2004). The likelihood principle, by itself, is not sufficient to build a method of inference but should be regarded as a minimum requirement of any viable form of inference. This is a controversial point of view for anyone familiar with modern econometrics literature. Much of this literature is devoted to methods that do not obey the likelihood principle (Rossi, Allenby, and McCulloch, 2005).

Suppose $L(\theta\mid X)$ is the assumed likelihood function. Under MLE estimation, we would compute the mode (the maximal value of L, as a function of θ given the data X) 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 & Fernando, 1986).

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

$$\mathrm{P}(\theta\mid \mathrm{X}) = \frac{1}{\mathrm{P(X)}}\mathrm{P}(\mathrm{X}\mid\theta)\mathrm{P}(\theta) \tag{1.4.1}$$

$$= (\text{normalizing constant})\,\mathrm{P}(\mathrm{X}\mid\theta)\mathrm{P}(\theta) \tag{1.4.2}$$

$$= \text{constant .likelihood .prior} \tag{1.4.3}$$

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

$$P(\theta\mid X) \propto L(\theta\mid X_1)P(\theta) \tag{1.4.4}$$

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

$$P(X) = \int_{\theta} P(X \mid \theta)P(\theta)d\theta \tag{1.4.5}$$

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 $\mathrm{L}(\theta \mid \mathrm{X})\mathrm{d}\theta$ taken over the admissible range of θ is finite, then occasionally it will be convenient to refer to the quantity

$$\frac{\mathrm{l}(\theta \mid \mathrm{X})}{\int \mathrm{l}(\theta \mid \mathrm{X})\mathrm{d}\theta}$$

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 Prior Distribution and Some Important Types of Priors:

A prior distribution of a parameter is the probability that represents uncertainty about the parameter before the current data are examined. A random variable can be thought of as a variable that takes on a set of values with specified probability. In frequentist statistics, parameters are not repeatable random things but are fixed quantities, which mean that they cannot be considered as random variables. In contrast, in Bayesian statistics anything about which we are uncertain, including the true value of the parameter, can be thought of as being a random variable to which we can assign a probability distribution, known specifically as prior information. A fundamental feature of the Bayesian approach to statistics is the use of prior information in addition to the (sample) data. A proper Bayesian analysis will always incorporate genuine prior information, which will help to strengthen inferences about the true value of the parameter and ensure that any relevant information about it is not wasted.

Obviously, a critical feature of any Bayesian analysis is the use of prior. According to Diaconis and Ylvisaker (1985), there are three distinct Bayesian approaches for the selection of prior distributions. The classical Bayesian approach considers flat priors to represent objectivity in the analysis. The modern approach allows the priors to have characteristics like closure under sampling (conjugacy) (suggested by G.Barnard (1954) and later developed by Raiffa & Schlaifer (1961)) and specification of hyper parameter values according to some specific criteria. The third approach is followed by subjective Bayesians, depends on elicitation of prior distributions based on pre-existing scientific knowledge in the area of investigation.

Some standard approaches of priors are discussed in brief as:

i) Non-informative Priors: A prior distribution is non-informative if the prior is "flat" relative to the likelihood function. Such a prior is also known as "vague", "diffuse" priors. Thus, a prior $\mathrm{P}(\theta)$ is non-informative if it has minimal impact on the posterior distribution of θ. Many statisticians favor non-informative priors because they appear to be more objective. According to Jeffery (1983), non-informative priors provide a formal way of expressing ignorance of the value of the parameter over the permitted range.

ii) Informative prior: An informative prior is a prior that is not dominated by the likelihood and that has an impact on the posterior distribution. If a prior distribution dominates the likelihood, it is clearly an informative prior. On the other hand, the proper use of prior distributions illustrates the power of the Bayesian method: information gathered from the previous study, past experience, or expert opinion can be combined with current information in a natural way.

iii)Improper prior: A prior $\mathrm{P}(\theta)$ is said to be improper if $\int P(\theta)\, d\theta = \infty$. For example, a uniform prior distribution on the real line, $\mathrm{P}(\theta) \propto 1$, for $-\infty < \theta < \infty$, is an improper prior. Improper priors are often used in Bayesian inference since they usually yield non-informative priors and proper posterior distributions. Improper prior distributions can lead to posterior impropriety (improper posterior distribution). To determine whether a posterior distribution is proper, you need to make sure that the normalizing constant $\int L(X \mid \theta) P(\theta)\, d\theta$ is finite for all x. If an improper prior distribution leads to an improper posterior distribution, inference based on the improper posterior distribution is invalid.

iv) Conjugate Priors: A prior is said to be a conjugate prior for a family of distributions if the prior and posterior distributions are from the same family, which means that the form of the posterior has the same distributional form as the prior distribution. For example, if the likelihood is binomial, $X \sim Bin(n, \theta)$, a conjugate prior on θ is the beta distribution; it follows that the posterior distribution of θ is also a beta distribution. Other commonly used conjugate prior/likelihood combinations include the normal/normal, gamma/Poisson, gamma/gamma, and gamma/beta cases. The development of conjugate priors was partially driven by a desire for computational convenience—conjugacy provides a practical way to obtain the posterior distributions.

v) Jefferys' Prior: A very useful prior is Jefferys' prior (1961). It satisfies the local uniformity property: a prior that does not change much over the region in which the likelihood is significant and does not assume large values outside that range. It is based on the Fisher information matrix. Jeffrey's prior is defined as

$$P(\theta) \propto |I(\theta)|^{-1/2}$$

Where $\mathrm{I}(\theta)$ denotes the Fisher information matrix based on the likelihood function

$$\mathrm{L}(\mathrm{X} \mid \theta) : I(\theta) = -\left[E\left\{ \frac{\partial^2 \log L(X \mid \theta)}{\partial \theta^2} \right\} \right]$$

Jeffrey's prior is locally uniform and hence non-informative. It provides an automated scheme for finding a non-informative prior for any parametric model $L(X \mid \theta)$. Another appealing property of Jeffreys' prior is that it is invariant with respect to one-to-one transformations. The invariance property means that if you have a locally uniform prior on θ and $\phi(\theta)$ is a one-to-one function of θ, then $P(\phi(\theta)) = P(\theta).|\phi'(\theta)|^{-1}$ is a locally uniform prior for $\phi(\theta)$. This invariance principle carries through to multidimensional parameters as well. While Jeffreys' prior provides a general recipe for obtaining non-informative priors, it has some shortcomings: the prior is improper for many models, and it can lead to improper posterior in some cases; and the prior can be cumbersome to use in high dimensions.

1.6 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 some unknown parameter θ. That is, we repeat the experiment n independent times, observe the sample $x_1, x_2, ..., x_n$. The function of $x_1, x_2, ..., x_n$ use to estimate θ; say the statistic $U(x_1, x_2, ..., x_n)$ called an estimator of θ. We want it to be such that the computed estimate $U(x_1, x_2, ..., x_n)$ is usually close to θ.

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

1.6.1 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.

Here we shall discuss only maximum likelihood estimate and Bayesian estimation.

1.6.2 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 Prof.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's. 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 $x_1, x_2, ..., x_n$ we can write in the discrete case.

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = f(x_1, x_2, ..., x_n)$$

which is just the value of joint probability distribution of the random values $x_1, x_2, \ldots, x_n$ at the sample point $x_1, x_2, \ldots, x_n$ since the sample values has been observed and are therefore fixed numbers, we regard $f(x_1, x_{2,} \ldots, x_n; \theta)$ 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 $f(x_1, x_{2,} \ldots, x_n; \theta)$ is the value of joint pdf at the sample point $x_1, x_2, ..., x_n$ i.e.; the likelihood function at the sample value $x_1, x_2, \ldots, x_n$

$$L = \prod_{i=1}^{n} f(x_i, \theta)$$

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 $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$ 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 $\hat{\theta}$ is the solution if and only if

$$\frac{\partial L}{\partial \theta} = 0 and \frac{\partial^2 L}{\partial \theta^2} < 0$$

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

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0$$

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

1.6.3 Bayesian method of estimation:

Bayesian analysis synthesis two sources of information about the unknown parameters of interest. The first of these is the sample data, expressed formally by the likelihood function. The second is the prior distribution, which represents additional information that is available to investigator. Suppose we have a random sample of size n say $\mathrm{x}_1, \mathrm{x}_{2,} ...., \mathrm{x}_\mathrm{n}$ which we regard as independent identically distributed random variables with distribution function $(df) F(X \mid \theta)$ and pdf $\mathrm{f(x} \mid \theta)$ and where

θ a labeling parameter, real valued or a 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 a parameter. 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 items from a random sample of size n from a population with pdf $f(x|\theta)$ 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

$$P(\theta | x_1, x_{2,} ...., x_n) = \left\{ \prod_{i=1}^{n} f(x_i | \theta) \right\} = L(x_1, x_{2,} ...., x_n | \theta)$$

The marginal pdf of ($x_1, x_{2,} ...., x_n$) is given by

$$p(x_1, x_{2,} ...., x_n) = \int_{\Omega} p(x_1, x_{2,} ...., x_n | \theta) d\theta$$

And the conditional pdf of θ given data ($x_1, x_{2,} ...., x_n$) is given by

$$P(\theta | x_1, x_{2,} ...., x_n) = \frac{p(x_1, x_{2,} ...., x_n | \theta)}{p(x_1, x_{2,} ...., x_n)}$$

$$P(\theta | x_1, x_{2,} ...., x_n) = \frac{L(x_1, x_{2,} ...., x_n | \theta) p(\theta)}{\int_{\Omega} L(x_1, x_{2,} ...., x_n | \theta) p(\theta) d\theta}$$

Thus, prior to obtaining ($x_1, x_{2,} ...., x_n$) the variations in θ where represented by $P(\theta)$, known as prior distribution on θ however, after the data ($x_1, x_{2,} ...., x_n$) has been obtained in the light of the new information, the variation in θ are represented by $P(\theta | x_1, x_{2,} ...., x_n)$ 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 $P(\theta | x_1, x_{2,} ...., x_n)$ this process is the straight forward application pdf the Bayes theorem. Once the posterior distribution has been obtained it becomes the main object of study.

1.7 Marginal and Conditional inferences:

Often only a subset of 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, if θ is partitioned as (θ_1, θ_2), with θ_1 a p dimensional vector and θ_2 as (k-p) dimensional vector, then the marginal posterior density for θ_1 is given by

$$P(\theta_1 \mid x) = \frac{\int_{R_2} P(x \mid \theta) P(\theta) d\theta_2}{\int_R P(x \mid \theta) P(\theta) d\theta}$$

(1.7.1)

Similarly, the marginal posterior density for θ_2 is given by

$$P(\theta_2 \mid x) = \frac{\int_{R_1} P(x \mid \theta) P(\theta) d\theta_1}{\int_R P(x \mid \theta) P(\theta) d\theta} \qquad (1.7.2)$$

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

Conditional inferences for θ_1 given θ_2 ; and θ_2 given θ_1 can also be made using the posteriors

$$P(\theta_1 \mid x, \theta_2) = \frac{P(x \mid \theta_1, \theta_2) P(\theta_1 \mid \theta_2)}{\int_{R_1} P(x \mid \theta_1, \theta_2) P(\theta_1 \mid \theta_2) d\theta_1} \qquad (1.7.3)$$

and $$P(\theta_2 \mid x, \theta_1) = \frac{P(x \mid \theta_1, \theta_2) P(\theta_2 \mid \theta_1)}{\int_{R_2} P(x \mid \theta_1, \theta_2) P(\theta_2 \mid \theta_1) d\theta_2} \qquad (1.7.4)$$

Marginal and conditional inferences procedures are two entirely different things. In the former, we ignore one of the components of θ by integrating it out from the joint posterior $P(\theta \mid x)$, while in the later we control (or adjust) one of the components of θ (e.g., Khan 1997).

1.8 Predictive Distribution:

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

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

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

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

1.9 Methods of Posterior Modes:

Asymptotic normality of the posterior is the basic tool of large sample Bayesian inference. Under certain regularity conditions, in particular, if the likelihood is a continuous function of θ and that the maximum likelihood estimate, $\hat{\theta}$ of θ is not the boundary of the parameter space, the unimodal and almost symmetric posterior distribution of θ approaches normality with mean $\hat{\theta}$ and precision $I(\hat{\theta})$, Fisher Information evaluated at $\hat{\theta}$,for large sample sizes. It may be noted that for large samples, the likelihood dominates the prior distribution and, therefore the knowledge of likelihood is enough to obtain the normal approximation. Gelman et.al. (1995) give a number of counter examples to illustrate limitations of the large sample approximation to the posterior distribution. The Bayesian approach to parametric inference is conceptually simple and probabilistically elegant. However its numerical implication is not convenient since the posterior distributions are available as complicated functions. Although these approximations provide useful results in applications, neither gives any account for the cases when the mode is at boundary.

In the development of new simulation techniques, Laplace's method uses asymptotic arguments. Laplace's method is easier to implement and thus faster than the Monte Carlo methods , such as Gibbs sampling(Gelfand and Smith 1990), which requires a large number of simulations from the conditional densities. Laplace approximations to marginal densities and expectations can provide further insights to the problem at hand.

1.9.1 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

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

where $$I\left(\hat{\theta}\right) = -\frac{\partial^2 \log P(\theta \mid y)}{\partial \theta^2} \qquad (1.9.1)$$

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.

1.9.2 Laplace's Approximation:

Laplace's method is a family of asymptotic methods used to approximate integrals presented as a potential candidate for the tool box of techniques used for knowledge acquisition and probabilistic inference in belief networks with continuous variables. The method is promising for computing approximation for Bayes factor for use in the context of model selection, model uncertainty and mixtures of pdf's. It is 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

$$P(\theta \mid X) = (2\pi)^{\frac{-k}{2}} \left| I(\hat{\theta}) \right|^{\frac{1}{2}} \exp\left[\log P(\hat{\theta} \mid x)\right] \left(1 + O(n^{-1})\right) \qquad (1.9.2a)$$

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.

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

(1.9.2b)

Approximation (1.9.2a) is the well known Laplace's approximation of integrals (e.g., Tierney and Kadane, 1986). Laplace's approximation (1.9.2b) 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).

1.10 Some Important Distributions:

i) Exponential distribution: Exponential distribution is widely used as model in the areas ranging from studies on the lifetimes of manufactured item (e.g., Davis, 1952; Epstein, 1958)to research involving survival or remission times in chronic diseases (e.g., Feigl and Zelen, 1965). Let X has an exponential distribution with parameter $\theta(\theta > 0)$ if its probability density function $f(x)$ is given by

$$f(x) = \lambda e^{-\lambda x}; \lambda, x > 0$$

The distribution is often written using the parameterization $\theta = \lambda^{-1}$, in which the pdf Becomes

$$f(x) = \frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right), x \geq 0$$

whose parameter θ is called rate parameter with mean θ and variance θ^2 respectively. The distribution where $\theta = 1$ is called the standard exponential distribution.

The most important properties of the exponential distribution is the memory less property i.e., probability of its surviving an additional h hours is exactly the same as the probability of surviving h hours of a new item.

$$P(X \leq (x + y \mid X > x)) = P(X \leq y)$$

where X is the time we need to wait before a certain events occurs. This property says that events happens during a time interval of length y is independent of how much time has already elapsed (x) without the event happening.

The pdf of two parameter exponential distribution is given by

$$f(x;\mu,\theta) = \frac{1}{\theta}\exp\left[\frac{-(x-\mu)}{\theta}\right], \qquad -\infty < \mu < x < \infty; \quad \theta > 0$$

ii) Gamma distribution: Gamma distribution has been quite extensively used as a lifetime model, though not censored. The gamma distribution is most widely used model for precipitation data. It fits a wide variety of lifetime data adequately, besides failure process models that lead to it. The gamma distribution has a pdf of the form

$$f(x;\alpha,\beta) = \frac{\alpha}{\Gamma(\beta)}\frac{(\alpha x)^{\beta-1}}{\alpha^{\beta}}\exp\{-\alpha x\}; x > 0; \alpha,\beta > 0$$

Where $\alpha, \beta > 0$ are parameters α^{-1} is a scale parameter and β is sometimes called the index or shape parameter. For $\beta = 1$, the gamma distribution reduces to the one parameter exponential distribution with parameters α has pdf

$$f(x) = \frac{1}{\alpha}\exp\left\{\frac{-x}{\alpha}\right\}; x > 0; \alpha > 0$$

For $\alpha = 1$ the distribution is called the one parameter gamma distribution and has pdf

$$f(x;\beta)=\frac{x^{\beta-1}}{\Gamma(\beta)}\exp\{-x\}\ ; x>0;\ \beta>0$$

The incomplete gamma distribution is given by:-

$$f(x;\beta)=\frac{1}{\Gamma(\beta)}\int_0^{\beta}u^{k-1}e^{-u}du$$

The moments of gamma distribution can be obtained as

$$E(X)=\frac{\beta}{\alpha}\qquad and \qquad V(X)=\frac{\beta}{\alpha^2}$$

Gamma distribution does not fit a wide variety of lifetime data adequately, however, and there are failure process models that lead to it. It also arises in some situations involving the exponential distribution; because of the well known results that the sum of independently and identically distributed exponential random variables have a gamma distribution. The distribution is also written using the parameterization $\alpha=\lambda^{-1}$, in which the pdf becomes

$$f(x;\alpha,\beta)=\frac{1}{\alpha\Gamma(\beta)}\frac{(x)^{\beta-1}}{\alpha^{\beta}}\exp\left\{\frac{-x}{\alpha}\right\}\ ;x>0$$

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

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

with mean μ and variance σ^2 , $\sigma>0$.

The normal distribution curve is bell shape and symmetrical about the line $x=\mu$. The mode and medium of the normal curve lies at the point $x=\mu$. The area under the normal curve within its range $-\infty to \infty$ in always unity i.e. $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}dx=1$. One of the greatest reasons behind the extensive use & application of normal distribution lies in central limit theorem which states: If $x_1,x_2,...,x_n$ is a random sample of size n from any population with mean μ and variance σ^2. The distribution of sample mean $\bar{x}$ 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

POSTERIOR APPROXIMATIONS TO EXPONENTIAL DISTRIBUTION

2.1 Introduction

The exponential distribution occupies an important position in the analysis of data. In probability theory and statistics, the exponential distribution is a family of continuous probability distribution. Historically, the exponential distribution was the first lifetime model for which statistical methods were extensively developed. It describes the time between events in the Poisson process i.e., a process in which events occur continuously and independently at a constant rate. Work by Sukhatmi (1937), Epstein and Sobel (1953, 1954, 1955) and Epstein (1954, 1960a) Bartholomew (1957), gave numerous results and popularized the exponential as a lifetime distribution, especially in the area of industrial life testing. Many authors have contributed to the statistical methodology of the distribution. The lengthy bibliographies of Mendenhall (1958), Govindarajulu (1964), Johnson and Kotz (1970), Johnson, Kotz and Balakrishnan (1994, 1995) and Lawless (2003), Ahmad (2006), Ahmed et. al. (2007 & 2010), contains a large number of papers in this area.

A random variable X has an exponential distribution with parameter $\theta(\theta > 0)$ if its probability density function $f(x)$ is of the form

$$f(x) = \frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right), x \geq 0; \theta > 0 \tag{2.1.1}$$

with mean θ and variance θ^2 respectively.

2.2 Maximum Likelihood Estimator of Exponential Distribution:

Let $X = x_1, x_2, ..., x_n$ be a random sample of size n with probability density function given as

$$\mathbf{f(x) = \frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right)} \tag{2.2.1}$$

The likelihood function is given as

$$L(x \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n}\exp\left(\frac{-n\bar{x}}{\theta}\right)$$

Applying log on both sides we get

$$LogL(x \mid \theta) = -n\log\theta - \frac{n\bar{x}}{\theta} \tag{2.2.2}$$

Differentiating (2.2.2) w.r.t θ, and equating to zero,

$$\frac{\partial \log L(x \mid \theta)}{\partial \theta} = 0$$

$$\frac{-n}{\theta}-\frac{n\bar{x}}{\theta^2}=0$$

$$\Rightarrow \hat{\theta}=\bar{x} \tag{2.2.3}$$

2.3 Bayesian Estimation for Exponential Distribution using Different Priors:

A detailed study on the Bayesian method of estimation has been done and presented in quite an interesting manner by Lindley (1965, 1971). Kale and Sinha (1980, 1983, and 1986) studied the Bayesian estimation of Exponential distribution in details. Bernardo and Smith (1994); Carlin and Levis (1996); Balakrishnan and Ma (1997); Viet (1986) and, Berger (1982, 1988), Ahmad (2006), Ahmed et. al (2007) & Ahmad & Bhat (2010) added more results to Bayesian estimation for Exponential distribution.

Sinha (1986) obtained the Bayes estimator of one parameter exponential distribution

$$f(x,\theta)=\frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right), x>0, \theta>0$$

If $f(\theta,x)$ is treated as a function of θ, then it will be likelihood of θ for single observation. A straight forward computation gives the Fisher information.

$$I(\theta)=-E\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]=\frac{1}{\theta^2}$$

Hence Jeffrey's prior $g(\theta)=\frac{1}{\theta}$, which is an improper (or quasi) prior since

$$\int_0^\infty g(\theta)\neq 1$$

Let us consider a more general class of priors,

$$g(\theta)\propto\left(\frac{1}{\theta^c}\right), \qquad c\geq 0$$

If $x_1, x_2, ..., x_n$ is a random sample from (2.2.1), then the likelihood function on this is given by

$$L(x\mid\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\frac{1}{\theta^n}\exp\left(-\frac{n\bar{x}}{\theta}\right)$$

Hence the posterior distribution of is given by

$$P(\theta\mid x)\propto\frac{1}{\theta^n}\exp\left(\frac{-n\bar{x}}{\theta}\right)\frac{1}{\theta^c}$$

$$P(\theta \mid x) = \frac{k}{\theta^{n+c}} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

(2.3.1)

where $$k^{-1} = \int_0^\infty \frac{1}{\theta^{n+c-1+1}} \exp\left(\frac{-n\bar{x}}{\theta}\right) d\theta$$

$$= \frac{\Gamma(n+c-1)}{(n\bar{x})^{n+c-1}}$$

or $$k = \frac{(n\bar{x})^{n+c-1}}{\Gamma(n+c-1)}$$

Thus

$$P(\theta \mid x) = \frac{(n\bar{x})^{n+c-1}}{\Gamma(n+c-1)\theta^{n+c}} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

(2.3.2)

Bayes estimator of θ is given by

$$\hat{\theta} = E(x \mid \theta) = \int_0^\infty \theta \; P(\theta \mid x) d\theta$$

$$= \int_0^\infty \theta \frac{(n\bar{x})^{n+c-1}}{\Gamma(n+c-1)\theta^{n+c}} \exp\left(\frac{-n\bar{x}}{\theta}\right) d\theta$$

$$= \frac{(n\bar{x})^{n+c-1}}{\Gamma(n+c-1)} \int_0^\infty \frac{\exp\left(\frac{-n\bar{x}}{\theta}\right)}{\theta^{n+c-1}} d\theta$$

$$= \frac{(n\bar{x})^{n+c-1}}{\Gamma(n+c-1)} \frac{\Gamma(n+c-2)}{(n\bar{x})^{n+c-2}}$$

$$= \frac{(n\bar{x})}{(n+c-2)}$$

(2.3.3)

For c=0, $g(\theta)=1$, we have

$$\hat{\theta} = \frac{(n\bar{x})}{(n-2)}$$

For c=1, $g(\theta) = \frac{1}{\theta}$,we have

$$\hat{\theta} = \frac{(n\bar{x})}{(n-1)}$$

For c=2, $g(\theta) = \frac{1}{\theta^2}$, we have

$$\hat{\theta} = \bar{x}$$

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

For c=3, $g(\theta)=\frac{1}{\theta^3}$ we have

$$\hat{\theta}=\frac{(n\bar{x})}{(n+1)}$$

This is well known minimum mean- square error (MSE) of θ.

Consider another class of priors given by

$$g(\theta)\propto\frac{\exp\left(\frac{-a}{\theta}\right)}{\theta^c};0<\theta<\infty;a,c\geq 0$$

$g(\theta)$ is a proper prior for c>0. The posterior of θ is given by

$$P(\theta\mid x)=\frac{(n\bar{x}+a)^{n+c-1}}{\Gamma(n+c-1)\theta^{n+c}}\exp\left(\frac{-(n\bar{x}+a)}{\theta}\right)$$

and Bayes estimate is given by

$$\hat{\theta}=\frac{n\bar{x}+a}{n+c-2}$$

Note that by putting a=0, we get the results obtained earlier in (2.3.3).

We now consider an inverted Gamma Prior (Raffier & Schlaifer, 1961) as the prior distribution of θ. Such a prior is given by

$$g(\theta)=\frac{1}{a\Gamma b}\exp\left(\frac{-a}{\theta}\right)\left(\frac{a}{\theta}\right)^{b+1};0<\theta<\infty;a,b>0$$

The Posterior distribution for θ is given by

$$P(\theta\mid x)\propto\frac{1}{\theta^n}\exp\left(\frac{-n\bar{x}}{\theta}\right)\frac{1}{a\Gamma b}\exp\left(\frac{-a}{\theta}\right)\left(\frac{a}{\theta}\right)^{b+1}$$

$$P(\theta\mid x)=\frac{k}{\theta^{n+b+1}}\exp\left(\frac{-(n\bar{x}+a)}{\theta}\right)$$

(2.3.4)

where $k^{-1}=\int_0^\infty\frac{1}{\theta^{n+b+1}}\exp\left(\frac{-(n\bar{x}+a)}{\theta}\right)d\theta$

$$k^{-1}=\frac{\Gamma(n+b)}{(n\bar{x}+a)^{n+b}}$$

or $$k=\frac{(n\bar{x}+a)^{n+b}}{\Gamma(n+b)}$$

$\therefore$ $$P(\theta \mid x) = \frac{(n\bar{x}+a)^{n+b}}{\Gamma(n+b)\theta^{n+b+1}} \mathrm{e} \quad \left(\frac{-(n\bar{x}+a)}{\theta}\right)$$

(2.3.5)

Thus Bayes estimator of θ is given by

$$\hat{\theta} = E(x \mid \theta) = \int_0^\infty \theta \, P(\theta \mid x) d\theta$$

$$= \int_0^\infty \theta \frac{(n\bar{x}+a)^{n+b}}{\Gamma(n+b)\theta^{n+b+1}} \exp\left(\frac{-(n\bar{x}+a)}{\theta}\right) d\theta$$

$$= \frac{(n\bar{x}+a)^{n+b}}{\Gamma(n+b)} \int_0^\infty \frac{1}{\theta^{n+b}} \exp\left(\frac{-(n\bar{x}+a)}{\theta}\right) d\theta$$

$$= \frac{(n\bar{x}+a)^{n+b}}{\Gamma(n+b)} \frac{\Gamma(n+b-1)}{(n\bar{x}+a)^{n+b-1}}$$

$$\hat{\theta} = \frac{n\bar{x}+a}{n+b-1} \qquad (2.3.6)$$

This is the same estimator considered before with b=c=1.

2.4 Laplace's Approximation for Exponential distribution:

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.

The likelihood function of exponential distribution is

$$L(x \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

Now we consider a more general class of priors,

$$\mathrm{g}(\theta) = \frac{1}{\theta^{\mathrm{c}}}, \mathrm{c} > 0$$

and hence, the posterior is

$$P(\theta \mid x) \propto L(x \mid \theta) g(\theta)$$

$$P(\theta \mid x) \propto \frac{1}{\theta^n} \exp\left(\frac{-n\bar{x}}{\theta}\right) \frac{1}{\theta^c}$$

$$P(\theta \mid x) \propto \frac{1}{\theta^{n+c}} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

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

$$\log P(\theta \mid x) = constant - (n+c)\log\theta - \frac{n\bar{x}}{\theta}$$

The posterior mode of this density is readily obtained as

$$\hat{\theta} = \frac{(n\bar{x})}{(n+c)} \tag{2.4.1}$$

The second derivative of the log-posterior density $\log P(\theta \mid x)$ at mode $\hat{\theta}$ is $-\frac{(n+c)^3}{(n\bar{x})^2}$

$$\therefore \quad I(\hat{\theta}) = \frac{(n+c)^3}{n\bar{x}^2} \tag{2.4.2}$$

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

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

$$\approx (2\pi)^{\frac{-1}{2}} \frac{(n+c)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} - (n+c)\log\theta + \frac{n\bar{x}}{\hat{\theta}} + (n+c)\log\hat{\theta}\right]$$

$$\approx (2\pi)^{\frac{-1}{2}} \frac{(n+c)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} - (n+c)\log\theta + n + c + (n+c)\log\left(\frac{n\bar{x}}{n+c}\right)\right]$$

$$\approx (2\pi)^{\frac{-1}{2}} \frac{(n+c)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n + c\right] \exp\left[-(n+c)\log\theta + (n+c)\log\left(\frac{n\bar{x}}{n+c}\right)\right]$$

$$\approx (2\pi)^{\frac{-1}{2}} \frac{(n+c)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n + c\right] \exp\left[\log\left(\frac{n\bar{x}}{\theta(n+c)}\right)^{n+c}\right] \tag{2.4.3}$$

For c=0, $g(\theta) = 1$ (2.4.3) becomes

$$P(\theta \mid X) \approx (2\pi)^{\frac{-1}{2}} \frac{(n)^{1/2}}{\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n\right] \exp\left[\log\left(\frac{\bar{x}}{\theta}\right)^{n}\right]$$

For c=1, $g(\theta) = 1/\theta$ (2.4.3) becomes

$$P(\theta \mid X) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+1)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n + 1\right] \exp\left[\log\left(\frac{n\bar{x}}{\theta(n+1)}\right)^{n+1}\right]$$

For c=2, $g(\theta) = 1/\theta^2$, (2.4.3) we have

$$P(\theta \mid X) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+2)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n + 2\right] \exp\left[\log\left(\frac{n\bar{x}}{\theta(n+2)}\right)^{n+2}\right]$$

For c=3, $g(\theta) = 1/\theta^3$, we have

$$P(\theta \mid X) \approx (2\pi)^{\frac{-1}{2}} \frac{(n+3)^{3/2}}{n\bar{x}} \exp\left[\frac{-n\bar{x}}{\theta} + n + 3\right] \exp\left[\log\left(\frac{n\bar{x}}{\theta(n+3)}\right)^{n+3}\right]$$

Laplace approximation is also discussed by Ahmad (2006) & Ahmad et.al. (2007).

2.5 Normal Approximation for Exponential distribution:

Let $x_1, x_2, \ldots\ldots, x_n$ be iid observations from an exponential distribution

$$f(x, \theta) = \frac{1}{\theta} \exp\left(\frac{-x}{\theta}\right), x > 0, \theta > 0$$

The likelihood function is given by

$$L(x \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

We consider a more general class of priors,

$$g(\theta) = \frac{1}{\theta^c}, c \geq 0$$

The posterior distribution is given by

$$P(\theta \mid x) \propto \frac{1}{\theta^n} \exp\left(\frac{-n\bar{x}}{\theta}\right) \frac{1}{\theta^c}$$

$$P(\theta \mid x) \propto \frac{1}{\theta^{n+c}} \exp\left(\frac{-n\bar{x}}{\theta}\right)$$

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

$$\log P(\theta \mid x) = cons\tan t - (n+c)\log\theta - \frac{n\bar{x}}{\theta}$$

The first derivative is

$$\frac{\partial}{\partial\theta} \log P(\theta \mid x) = \frac{-n+c}{\theta} - \frac{n\bar{x}}{\theta^2}$$

from which the posterior mode is readily obtained as

$$\hat{\theta} = \frac{(n\bar{x})}{(n+c)} \tag{2.5.1}$$

The second derivative of the log-posterior density is

$$\frac{\partial^2 \log P(\theta \mid x)}{\partial\theta^2} = \frac{n+c}{\theta^2} - \frac{2n\bar{x}}{\theta^3}$$

and hence, negative of Hessian is

$$I(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2} \log P(\theta \mid x) = \frac{2n\bar{x}}{\theta^3} - \frac{n+c}{\theta^2}$$

$$\therefore \quad I(\hat{\theta}) = \frac{(n+c)^3}{(n\bar{x})^2} \tag{2.5.2}$$

Thus, the posterior distribution can be approximated as

$$P(\theta \mid x) \sim N\left(\frac{n\bar{x}}{n+c}, \frac{(n\bar{x})^2}{(n+c)^3}\right) \tag{2.5.3}$$

For c=0, $g(\theta)=1$ (uniform prior), we have

$$P(\theta \mid x) \sim N\left(\bar{x}, \frac{\bar{x}^2}{n}\right)$$

For c=1, $g(\theta)=1/\theta$ (Jeffrey's prior), we have

$$P(\theta \mid x) \sim N\left(\frac{n\bar{x}}{n+1}, \frac{(n\bar{x})^2}{(n+1)^3}\right)$$

For c=2, $g(\theta)=1/\theta^2$, we have

$$P(\theta \mid x) \sim N\left(\frac{n\bar{x}}{n+2}, \frac{(n\bar{x})^2}{(n+2)^3}\right)$$

For c=3, $g(\theta)=1/\theta^3$, we have

$$P(\theta \mid x) \sim N\left(\frac{n\bar{x}}{n+3}, \frac{(n\bar{x})^2}{(n+3)^3}\right)$$

Now, consider another class of priors given by

$$g(\theta) \propto \frac{\exp\left(-\frac{a}{\theta}\right)}{\theta^c} \quad , \quad c \geq 0,\ a \geq 0, \quad \theta > 0$$

The log-posterior density of θ is given by

$$\log P(\theta \mid x) = cons\tan t - \frac{n\bar{x}+a}{\theta} - (n+c)\log\theta$$

The first derivative is

$$\frac{\partial}{\partial\theta}\log P(\theta \mid x) = \frac{n\bar{x}+a}{\theta^2} - \frac{n+c}{\theta}$$

From which the posterior mode is readily obtained as

$$\hat{\theta} = \frac{n\bar{x}+a}{n+c} \tag{2.5.4}$$

The second derivative of the log posterior density is

$$\frac{\partial^2}{\partial\theta^2}\log P(\theta \mid x) = -\frac{2(n\bar{x}+a)}{\theta^3} + \frac{n+c}{\theta^2}$$

$$I(\theta) = -\frac{\partial^2 \log P(\theta \mid x)}{\partial\theta^2} = \frac{2(n\bar{x}+a)}{\theta^3} - \frac{n+c}{\theta^2}$$

The second derivative at the mode $\hat{\theta}$ is then

$$I(\hat{\theta}) = \frac{2(n\bar{x}+a)(n+c)^3}{(n\bar{x}+a)^3} - \frac{(n+c)(n+c)^2}{(n\bar{x}+a)^2}$$

$$I(\hat{\theta}) = \frac{(n+c)^3}{(n\bar{x}+a)^2} \qquad (2.5.5)$$

$$[I(\hat{\theta})]^{-1} = \frac{(n\bar{x}+a)^2}{(n+c)^3}$$

Thus, the posterior distribution can be approximated as

$$P(\theta \mid x) \sim N\left(\frac{n\bar{x}+a}{n+c}, \frac{(n\bar{x}+a)^2}{(n+c)^3}\right)$$

(2.5.6)

By putting a=0, we have the results as obtained earlier in (2.5.3)

Now, consider an inverted gamma as the prior

$$g(\theta) = \frac{1}{a\Gamma b}\exp\left(\frac{-a}{\theta}\right)\left(\frac{a}{\theta}\right)^{b+1}, \theta > 0; a, b > 0$$

The posterior density of θ is given by

$$P(\theta \mid x) \propto \frac{1}{\theta^{n+b+1}}\exp\left(\frac{-(n\bar{x}+a)}{\theta}\right)$$

The log-posterior density of θ is given by

$$\log P(\theta \mid x) = constant - \frac{n\bar{x}+a}{\theta} - (n+b+1)\log\theta$$

The first derivative is

$$\frac{\partial \log P(\theta \mid x)}{\partial\theta} = \frac{n\bar{x}+a}{\theta^2} - \frac{n+b+1}{\theta}$$

From which the posterior mode is readily obtained as

$$\hat{\theta} = \frac{n\bar{x}+a}{n+b+1} \qquad (2.5.7)$$

The second derivative of the log-posterior density is

$$\frac{\partial^2 \log P(\theta \mid x)}{\partial\theta^2} = -\frac{2(n\bar{x}+a)}{\theta^3} + \frac{n+b+1}{\theta^2}$$

$$I(\theta) = -\frac{\partial^2 \log P(\theta \mid x)}{\partial\theta^2} = \frac{2(n\bar{x}+a)}{\theta^3} - \frac{n+b+1}{\theta^2}$$

The second derivative at the mode $\hat{\theta}$ is then

$$I(\hat{\theta}) = \frac{2(n\bar{x}+a)(n+b+1)^3}{(n\bar{x}+a)^3} - \frac{(n+b+1)(n+b+1)^2}{(n\bar{x}+a)}$$

$$I(\hat{\theta}) = \frac{(n+b+1)^3}{(n\bar{x}+a)^2} \qquad (2.5.8)$$

$$\therefore \qquad \left[I(\hat{\theta})\right]^{-1} = \frac{(n\bar{x}+a)^2}{(n+b+1)^3}$$

Thus, the posterior distribution can be approximated as

$$P(\theta \mid x) \approx N\left(\frac{n\bar{x}+a}{n+b+1}, \frac{(n\bar{x}+a)^2}{(n+b+1)^3}\right)$$

(2.5.9)

For b=c-1, the result is same as in (2.5.3).

2.6 Bayesian Estimation for Two Parameter Exponential Distribution:

The pdf of two parameter exponential distribution is given by

$$f(x;\mu,\theta) = \frac{1}{\theta}\exp\left[\frac{-(x-\mu)}{\theta}\right], \qquad -\infty < \mu < x < \infty; \quad \theta > 0$$

(2.6.1)

The likelihood function is given by

$$\begin{aligned} L(x \mid \theta) &= \frac{1}{\theta^n}\exp\left[\frac{-1}{\theta}\sum(x_i-\mu)\right] \\ &= \frac{1}{\theta^n}\exp\left[\frac{-1}{\theta}\sum\left\{(x_i - x_{(1)}) + (x_{(1)}-\mu)\right\}\right] \\ &= \frac{1}{\theta^n}\exp\left[\frac{-1}{\theta}\left\{s + n(x_{(1)}-\mu)\right\}\right] \qquad (2.6.2) \end{aligned}$$

where $x_{(1)}$ is the first order statistic in the sample $X = (x_1, x_2, \ldots, x_n)$ such that $x_{(1)} \le x_{(2)} \le \ldots.. \le x_{(n)}$ and $S = \sum_{i=1}^{n}(x_i - x_{(1)})$.

Bayesian estimator of θ & μ under Jeffery's prior

$$\mathbf{g(\mu,\theta) \propto \frac{1}{\theta^c}; c > 0}$$

Thus joint posterior distribution of θ & μ is given as

$$\begin{aligned} P(\theta,\mu \mid X) &\propto L(x \mid \theta) \quad g(\mu,\theta) \\ &= \frac{k}{\theta^{n+c}}\exp\left[\frac{-1}{\theta}\left\{s + n(x_{(1)}-\mu\right\}\right] \qquad (2.6.3) \end{aligned}$$

Where K is a normalizing constant and is given by

$$K^{-1} = \int_{-\infty}^{x_{(1)}}\int_{0}^{\infty}\frac{1}{\theta^{n+c}}\exp\left[\frac{-1}{\theta}\left\{s + n(x_{(1)}-\mu)\right\}\right]d\mu\, d\theta$$

$$K^{-1} = \int_{-\infty}^{x_{(1)}}\left[\int_{0}^{\infty}\frac{-1}{\theta^{n+c}}\exp\left[\frac{-1}{\theta}\left\{s + n(x_{(1)}-\mu)\right\}\right]d\theta\right]d\mu$$

$$K^{-1} = \int_{-\infty}^{x_{(1)}} \frac{\Gamma(n+c-1)}{[s+n(x_{(1)}-\mu)]^{n+c-1}}\, d\mu$$

$put \qquad \{s+n(x_{(1)}-\mu)\} = \nu; \quad \Rightarrow d\mu = \frac{-1}{n} d\nu$

$as \qquad \mu \to -\infty \Rightarrow \nu \to \infty;\ \mu \to x_{(1)} \Rightarrow \nu \to s$

$$\therefore \qquad K^{-1} = \frac{\Gamma(n+c-1)}{-n} \int_{\infty}^{s} \nu^{-(n+c-1)} d\nu$$

$$K^{-1} = \frac{-\Gamma n}{n}\left[\frac{\nu^{-(n+c-1)+1}}{-(n+c-1)+1}\right]_{\infty}^{s} = \frac{\Gamma(n+c-1) s^{-(n+c-2)}}{n(n+c-2)} = \frac{\Gamma(n+c-1)}{ns^{(n+c-2)}}$$

$$\therefore \qquad K = \frac{n\, s^{(n+c-2)}}{\Gamma(n+c-1)}$$

$$\therefore \qquad P(\mu,\theta \mid X) = \frac{n S^{n+c-2}}{\Gamma(n+c-2)\ \theta^{n+c}} \exp\left[\frac{-1}{\theta}\{S+n(x_{(1)}-\mu)\}\right]$$

(2.6.4)

For $c=0, g(\mu,\theta)=1$ (uniform prior), (2.6.4) becomes

$$P(\mu,\theta \mid X) = \frac{n S^{n-2}}{\Gamma(n-2)\ \theta^{n}} \exp\left[\frac{-1}{\theta}\{S+n(x_{(1)}-\mu)\}\right]$$

For $c=1,\ g(\mu,\theta)=\frac{1}{\theta}$, (2.6.4) becomes

$$P(\mu,\theta \mid X) = \frac{n S^{n-1}}{\Gamma(n-1)\ \theta^{n+1}} \exp\left[\frac{-1}{\theta}\{S+n(x_{(1)}-\mu)\}\right]$$

For $c=2,\ g(\mu,\theta)=\frac{1}{\theta^2}$, (2.6.4) becomes

$$P(\mu,\theta \mid X) = \frac{n S^{n}}{\Gamma(n)\ \theta^{n+2}} \exp\left[\frac{-1}{\theta}\{S+n(x_{(1)}-\mu)\}\right]$$

2.7 Marginal Posterior densities for μ and θ:

The marginal posterior density of μ is given by

$$P(\mu \mid X) = \int_0^{\infty} P(\mu,\theta \mid X)\, d\theta$$

$$= \frac{n\ S^{n+c-2}}{\Gamma(n+c-2)} \int_0^{\infty} \frac{1}{\theta^{n+c}} \exp\left[\frac{-1}{\theta}\{S+n(x_{(1)}-\mu)\}\right] d\theta$$

$$= \frac{n\ S^{n+c-2}}{\Gamma(n+c-2)} \ \frac{\Gamma(n+c-1)}{\{S+n(x_{(1)}-\mu)\}^{(n+c-1)}}$$

$$= \frac{n(n+c-2)\ S^{n+c-2}}{\{S+n(x_{(1)}-\mu)\}^{(n+c-1)}} \qquad (2.7.1)$$

For $c=0,\ g(\mu,\theta)=1$, $(2.7.1)$ becomes $P(\mu\,|\,x)=\dfrac{n\,(n-2)\,S^{n-2}}{\{S+n(x_{(1)}-\mu)\}^{n-1}}$

For $c=1,\ g(\mu,\theta)=\dfrac{1}{\theta}$, $(2.7.1)$ becomes $P(\mu\,|\,x)=\dfrac{n\,(n-1)\,S^{n-1}}{\{S+n(x_{(1)}-\mu)\}^{n}}$

For $c=2,\ g(\mu,\theta)=\dfrac{1}{\theta^2}$, $(2.7.1)$ becomes $P(\mu\,|\,x)=\dfrac{n^2\,S^{n-2}}{\{S+n(x_{(1)}-\mu)\}^{n+1}}$

The marginal posterior density of θ is given by

$$P(\theta\,|\,X)=\int_{-\infty}^{x_{(1)}} P(\mu,\theta\,|\,x)\,d\mu$$

put $\{s+n(x_{(1)}-\mu)\}=\nu;\ \Rightarrow d\mu=\dfrac{-1}{n}d\nu$

as $\mu\to-\infty\Rightarrow\nu\to\infty;\ \mu\to x_{(1)}\Rightarrow\nu\to s$

$$\therefore\quad P(\theta\,|\,X)=\frac{n\,S^{n+c-2}}{\Gamma(n+c-2)}\,\frac{1}{\theta^{n+c}}\int_{\infty}^{s}\exp\left[\frac{-v}{\theta}\right]\left(\frac{-1}{n}\right)dv$$

$$=\frac{S^{n+c-2}\exp\left(\dfrac{-S}{\theta}\right)}{\Gamma(n+c-2)\theta^{n+c-1}}\qquad(2.7.2)$$

For $c=0, g(\theta)=1$, $(2.7.1)$becomes $P(\theta\,|\,x)=\dfrac{S^{n+c-2}\exp\left(\dfrac{-S}{\theta}\right)}{\Gamma(n-2)\theta^{n-1}}$

For $c=1, g(\theta)=\dfrac{1}{\theta}$, $(2.7.2)$ becomes $P(\theta\,|\,x)=\dfrac{S^{n+c-2}\exp\left(\dfrac{-S}{\theta}\right)}{\Gamma(n-1)\theta^{n+c}}$

For $c=2,\ g(\theta)=\dfrac{1}{\theta^2}$, $(2.7.2)$ becomes $P(\theta\,|\,x)=\dfrac{S^{n+c}\exp\left(\dfrac{-S}{\theta}\right)}{\Gamma(n)\theta^{n+1}}$

2.8 Posterior estimates of μ and θ :

The posterior estimate of μ is given by

$$\hat{\mu}=E(\mu\,|\,x)=\int_{-\infty}^{x_{(1)}}\mu\ P(\mu\,|\,x)\ d\mu$$

$$=n\ (n+c-2)\ S^{n+c-2}\int_{-\infty}^{x_{(1)}}\frac{\mu}{\{S+n(x_{(1)}-\mu)\}^{n+c-1}}\,d\mu$$

put $\{s+n(x_{(1)}-\mu)\}=\nu;\ \Rightarrow\mu=\dfrac{nx_{(1)}-v+s}{n}\Rightarrow d\mu=\dfrac{-1}{n}d\nu$

as $\mu\to-\infty\Rightarrow\nu\to\infty;\ \mu\to x_{(1)}\Rightarrow\nu\to s$

$$\therefore \qquad \hat{\mu} = n\,(n+c-2)\,S^{n+c-2}\int_{\infty}^{s}\frac{-(nx_{(1)}-v+s)}{n^2\{v\}^{n+c-1}}\,dv$$

$$=\frac{-(n+c-2)\,S^{n+c-2}}{n}\int_{\infty}^{s}(nx_{(1)}-v+s)\,v^{-(n+c-1)}\,dv$$

$$=\frac{-(n+c-2)\,S^{n+c-2}}{n}\left[nx_{(1)}\int_{\infty}^{s}v^{-(n+c-1)}dv-\int_{\infty}^{s}v^{-(n+c-1)+1}dv+s\int_{\infty}^{s}v^{-(n+c-1)}dv\right]$$

$$=\frac{-(n+c-2)\,S^{n+c-2}}{n}\left[\frac{nx_{(1)}S^{-(n+c-2)}}{-(n+c-2)}+\frac{S^{-(n+c-3)}}{n+c-3}-\frac{S^{-(n+c-3)}}{n+c-2}\right]$$

$$=x_{(1)}-\frac{S}{n(n+c-3)}$$

(2.8.1)

For $c=0, g(\theta)=1$, (2.8.1) becomes $\hat{\mu}=x_{(1)}-\dfrac{S}{n(n-3)}$

For $c=1, g(\theta)=\dfrac{1}{\theta}$, (2.8.1) becomes $\hat{\mu}=x_{(1)}-\dfrac{S}{n(n-2)}$

For $c=2, g(\theta)=\dfrac{1}{\theta^2}$, (2.8.1) becomes $\hat{\mu}=x_{(1)}-\dfrac{S}{n(n-1)}$

The Posterior estimates of θ is given by

$$\hat{\theta}=E(\theta\mid X)=\int_0^{\infty}\theta\,P(\theta\mid X)\,d\theta$$

$$=\frac{nS^{n+c-2}}{\Gamma(n+c-2)}\int_0^{\infty}\frac{e^{-\frac{S}{\theta}}}{\theta^{n+c-2}}\,d\theta$$

$$=\frac{S^{n+c-2}\Gamma(n+c-3)}{\Gamma(n+c-2)S^{n+c-3}}=\frac{S}{n+c-3}$$

(2.8.2)

For $c=0, g(\theta)=1$,(2.8.2) becomes $E(\theta\mid x)=\dfrac{S}{(n-3)}$

For $c=1, g(\theta)=\dfrac{1}{\theta}$, (2.8.2)becomes $E(\theta\mid x)=\dfrac{S}{(n-2)}$

For $c=2, g(\theta)=\dfrac{1}{\theta^2}$, (2.8.2)becomes $E(\theta\mid x)=\dfrac{S}{(n-1)}$

Example 2.1 (Deshpande and Puorhit 2005):

Following are the time in days between successive earthquakes worldwide. An earthquake is included in the data set if its magnitude was at least 7.5 on Richter scale, or if over 1000 people were killed. Recordings start on 16th of December 1902 and ends on 14th march 1997.There were 63 earthquakes recorded altogether, and so 62 waiting times.

840,157,145,44,33,121,150,280,434,736,584,887,263,1901,695,294,562,721,76,710,46,402,194,759,319,460,40,1336,335,1334,454,36,667,40,556,99,304,375,567,139,780,203,436,30,384,129,9,209,599,83,832,328,246,1617,638,937,735,38,365,92,82,220.

In order to find the Bayesian estimates for above example of univariate exponential distribution, we have developed the programmes in S-PLUS and R software and the results are given in tables 2.1 and 2.2.

```
# Bayes estimates with different priors.

# S-PLUS

 bayesexp.est<-function(x)

 {

 n<-length(x)

 C<-c(0,1,2,3)

 estimate<-(n*mean(x))/(n+C-2)

 return(estimate)

 }

time<-
c(840,157,145,44,33,121,150,280,434,736,584,887,263,1901,695,294,562
,721,76,710,46,402,194,756,319,460,40,1336,335,1334,454,36,667,40,55
6,99,304,375,567,139,780,203,436,30,384,129,9,209,599,83,832,328,246
,1617,638,937,735,38,365,92,82,220)

bayesexp.est(time) # To get the output.
```

Table: 2.1: Bayes estimates of Exponential distribution with different priors.

Prior	Bayes Estimate
1	451.40
1/theta	444.00
1/theta^2	436.84
1/theta^3	429.90

 # Posterior density of Exponential Distribution using different priors in R Software.

 # Prior=1.

```
time<-
c(840,157,145,44,33,121,150,280,434,736,584,887,263,1901,695,294,562
,721,76,710,46,402,194,756,319,460,40,1336,335,1334,454,36,667,40,55
6,99,304,375,567,139,780,203,436,30,384,129,9,209,599,83,832,328,246
,1617,638,937,735,38,365,92,82,220)

   pos.exp<-function(theta=200)

   {

   n<-length(time)

   C<-0

   pos<-(n+C)*log(theta)+sum(time)/theta

   return(pos)

   }

library(stats4)

fit<-mle(pos.exp)

summary(fit)

> summary(fit)

Maximum likelihood estimation

Call:

mle(minuslogl = pos.exp)

Coefficients:

      Estimate Std. Error

theta 436.5486   55.40446

-2 log L: 877.866
```

 # Prior=1/theta.

```
pos.exp<-function(theta=200)
   {
   n<-length(time)
   C<-1
   pos<-(n+C)*log(theta)+sum(time)/theta
   return(pos)
   }
library(stats4)
fit<-mle(pos.exp)
summary(fit)
Maximum likelihood estimation
Call:
mle(minuslogl = pos.exp)
Coefficients:
      Estimate Std. Error
theta 429.6785   54.10542
-2 log L: 890.009
```

 # Prior=1/theta^2

```
pos.exp<-function(theta=200)
   {
   n<-length(time)
   C<-2
   pos<-(n+C)*log(theta)+sum(time)/theta
   return(pos)
   }
library(stats4)
fit<-mle(pos.exp)
  summary(fit)
Maximum likelihood estimation
```

```
Call:

mle(minuslogl = pos.exp)

Coefficients:

      Estimate Std. Error

theta  423.012   52.85375

-2 log L: 902.1204
```

 # Prior=1/theta^3.

```
pos.exp<-function(theta=200)
   {
   n<-length(time)
   C<-3
   pos<-(n+C)*log(theta)+sum(time)/theta
   return(pos)
   }
library(stats4)
fit<-mle(pos.exp)
 Maximum likelihood estimation

Call:

mle(minuslogl = pos.exp)

Coefficients:

      Estimate Std. Error

theta 416.5415   51.64886

-2 log L: 914.2005  summary(fit)
```

Table 2.2: Posterior mode and Posterior standard error of Exponential distribution with different priors.

Prior	Posterior mode	Posterior Standard error
1	436.5486	55.40446
1/theta	429.6785	54.10542
1/theta^2	423.012	52.85375
1/theta^3	416.5415	51.64886

Comparing Normal Approximation of Exponential Distribution with different priors in S-PLUS and R.

```
Norm.app<-function(x)

{

n<-length(x)

theta<-seq(200,700,length=1500)

plot(theta,dnorm(theta,mean=mean(x),sd=sqrt((mean(x))^2/n)),

xlab="theta",ylab="p(theta|x)",ylim=c(0,0.008),

      main="Posterior Density for Time with different Priors",
      sub="Figure 2.1:Comparing Normal Approximation with different
      priors",type="l",col=3)

lines(theta,dnorm(theta,mean=(n*mean(x)/(n+1)),sd=sqrt((n*mean(x))^2
      /(n+1)^3)),col=4)

lines(theta,dnorm(theta,mean=(n*mean(x)/(n+2)),sd=sqrt((n*mean(x))^2
      /(n+2)^3)),col=5)

lines(theta,dnorm(theta,mean=(n*mean(x)/(n+3)),
      sd=sqrt((n*mean(x))^2/(n+3)^3)),col=6)

}

time<-
c(840,157,145,44,33,121,150,280,434,736,584,887,263,1901,695,294,562
,721,76,710,46,402,194,756,319,460,40,1336,335,1334,454,36,667,40,55
6,99,304,375,567,139,780,203,436,30,384,129,9,209,599,83,832,328,246
,1617,638,937,735,38,365,92,82,220)
```

```
Norm.app(time)

leg.names<-
c("Prior=1","Prior=1/theta","Prior=1/theta^2","Prior=1/theta^3")

legend(locator(1),leg.names,fill=3:6)
```

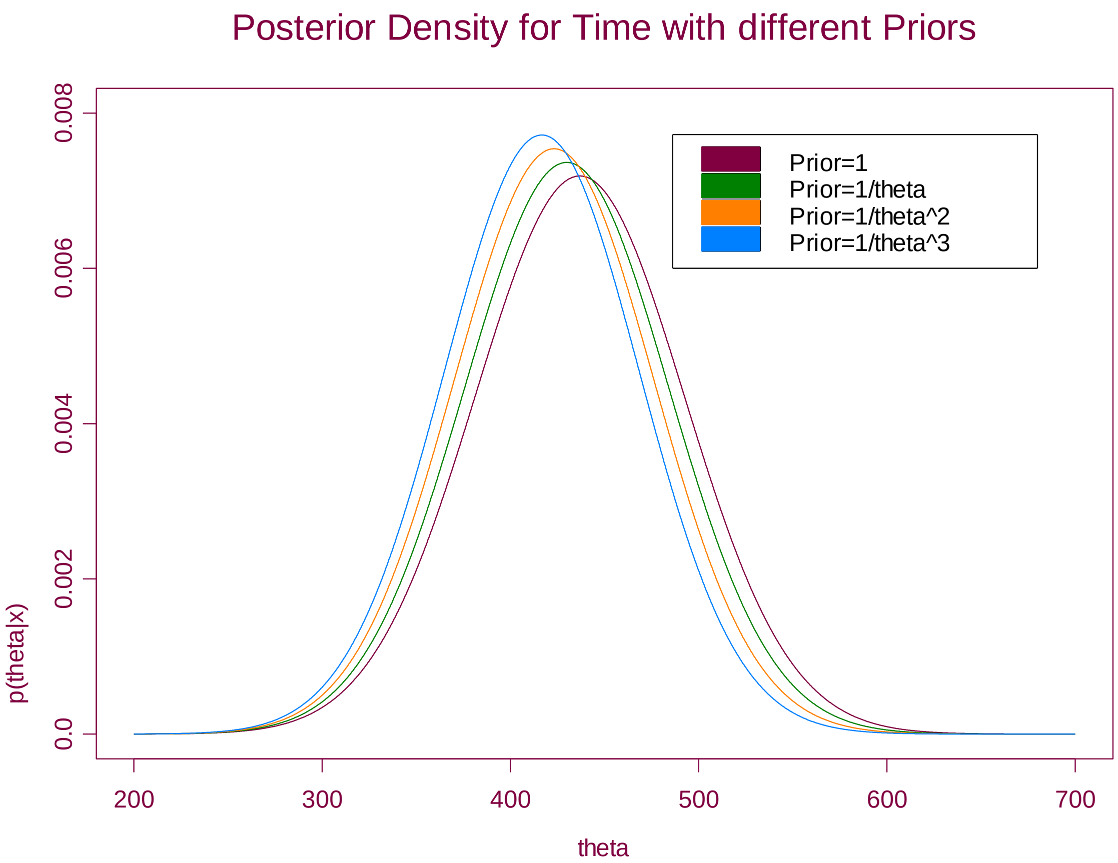


Figure 2.1:Comparing Normal Approximation with different priors

 # Comparing Laplace's Approximation of Exponential Distribution with different priors in S-PLUS and R .

```
Lap.app<-function(x)

{

n<-length(x)

theta<-seq(200,700,length=1500)

ptheta0<-(1/sqrt(2*pi))*(n^.5/mean(x))*
exp(sum(x)/theta+n)*exp(n*log(mean(x)/theta))

plot(theta,ptheta0, xlab="theta",ylab="p(theta|x)",ylim=c(0,0.008),

       main="Posterior Density for Time with Prior=1",
```

```
sub="Figure 2.2:Comparing Laplace's Approximation with different
priors", type="l", col=3)

ptheta1<-(1/sqrt(2*pi))*((n+1)^1.5)/(n*mean(x))*exp(-
         sum(x)/theta+n+1)* exp((n+1)*log(n*mean(x)/((n+1)*theta)))

lines(theta,ptheta1,col=4)

ptheta2<-(1/sqrt(2*pi))*((n+2)^1.5)/(n*mean(x))*exp(-
sum(x)/theta+n+2)*

           exp((n+2)*log(n*mean(x)/((n+2)*theta)))

lines(theta,ptheta2,col=5)

ptheta3<-(1/sqrt(2*pi))*((n+3)^1.5)/(n*mean(x))*

         exp(-
sum(x)/theta+n+3)*exp((n+3)*log(n*mean(x)/((n+3)*theta)))

lines(theta,ptheta3,col=6)

}

Lap.app(time)

leg.names<-
c("Prior=1","Prior=1/theta","Prior=1/theta^2","Prior=1/theta^3")

legend(locator(1),leg.names,fill=3:6)
```

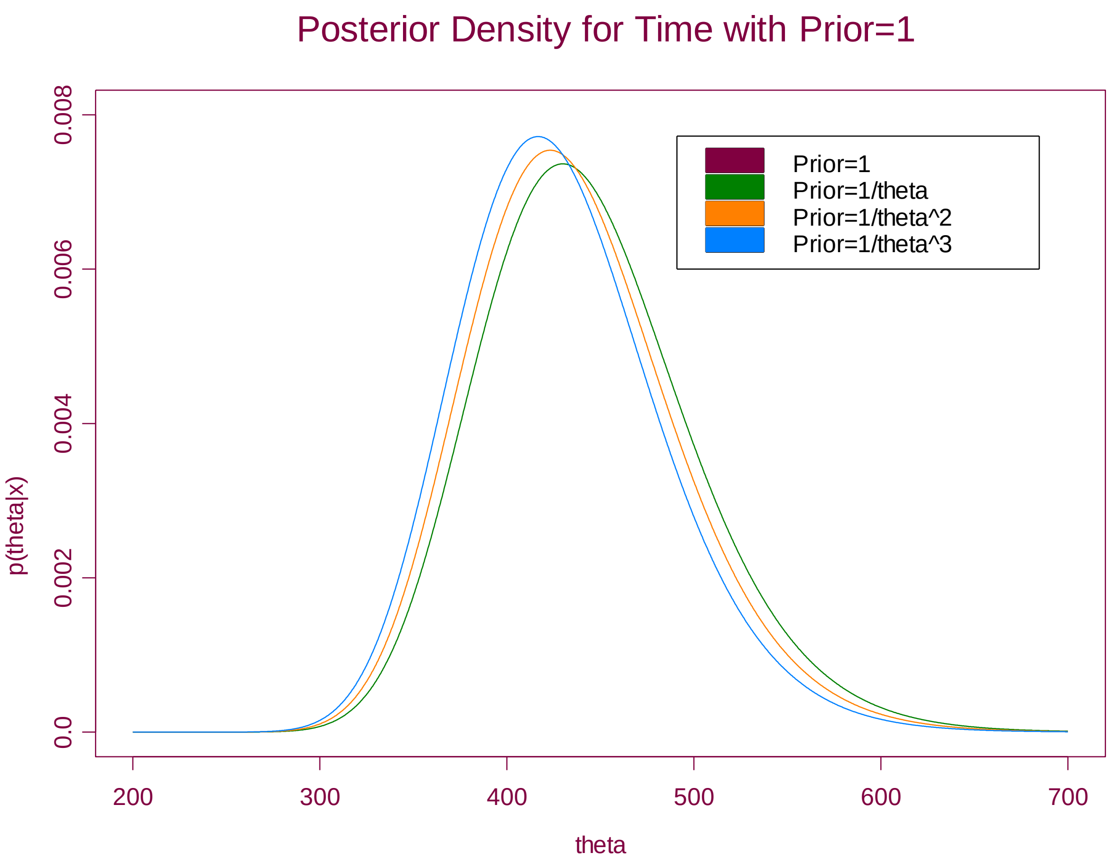


Figure 2.2:Comparing Laplace's Approximation with different priors

 # Comparison between Normal and Laplace's Approximation of Exponential distribution with different priors in S-PLUS and R .

```
Norm.lap<-function(x)

{

n<-length(x)

theta<-seq(200,700,length=800)

plot(theta,dnorm(theta,mean=mean(x),sd=sqrt((mean(x))^2/n)),xlab="th
eta",ylab="p(theta|x)",

      ylim=c(0,0.008),main="Posterior Density for Time with Prior=1",

      sub="Figure 2.3:Comparison between Normal and Laplace's
Approximation",type="l",col=3)

ptheta0<-(1/sqrt(2*pi))*(n^.5/mean(x))*exp(-
sum(x)/theta+n)*exp(n*log(mean(x)/theta))

lines(theta,ptheta0,col=4)
```

```
plot(theta,dnorm(theta,mean=(n*mean(x)/(n+1)),sd=sqrt((n*mean(x))^2/
(n+1)^3)),xlab="theta",

     ylab="p(theta|x)", ylim=c(0,0.008),main="Posterior Density for
Time with Prior=1/theta",

     sub="Figure 2.4: Comparison between Normal and Laplace,s
Approximation",type="l",col=3)

ptheta1<-(1/sqrt(2*pi))*((n+1)^1.5)/(n*mean(x))*exp(-
         sum(x)/theta+n+1)* exp((n+1)*log(n*mean(x)/((n+1)*theta)))

lines(theta,ptheta1,col=4)

plot(theta,dnorm(theta,mean=(n*mean(x)/(n+2)),sd=sqrt((n*mean(x))^2/
(n+2)^3)),xlab="theta", ylab="p(theta|x)",
ylim=c(0,0.008),main="Posterior Density for Time with
Prior=1/theta^2",

     sub="Figure 2.5: Comparison between Normal and Laplace's
Approximation",type="l",col=3)

ptheta2<-(1/sqrt(2*pi))*((n+2)^1.5)/(n*mean(x))*exp(-
sum(x)/theta+n+2)*

           exp((n+2)*log(n*mean(x)/((n+2)*theta)))

lines(theta,ptheta2,col=4)

plot(theta,dnorm(theta,mean=(n*mean(x)/(n+3)),sd=sqrt((n*mean(x))^2/
(n+3)^3)),xlab="theta",

ylab="p(theta|x)", ylim=c(0,0.008),main="Posterior Density for Time
with Prior=1/theta^3",

sub="Figure 2.6: Comparison between Normal and Laplace's
Approximation",type="l",col=3)

ptheta3<-(1/sqrt(2*pi))*((n+3)^1.5)/(n*mean(x))*

         exp(-
sum(x)/theta+n+3)*exp((n+3)*log(n*mean(x)/((n+3)*theta)))

lines(theta,ptheta3,col=4)

}

Norm.lap(time)
```

```
leg.names<-c("Normal Approximation","Laplace's Approximation")

legend(locator(1),leg.names,fill=3:4)
```

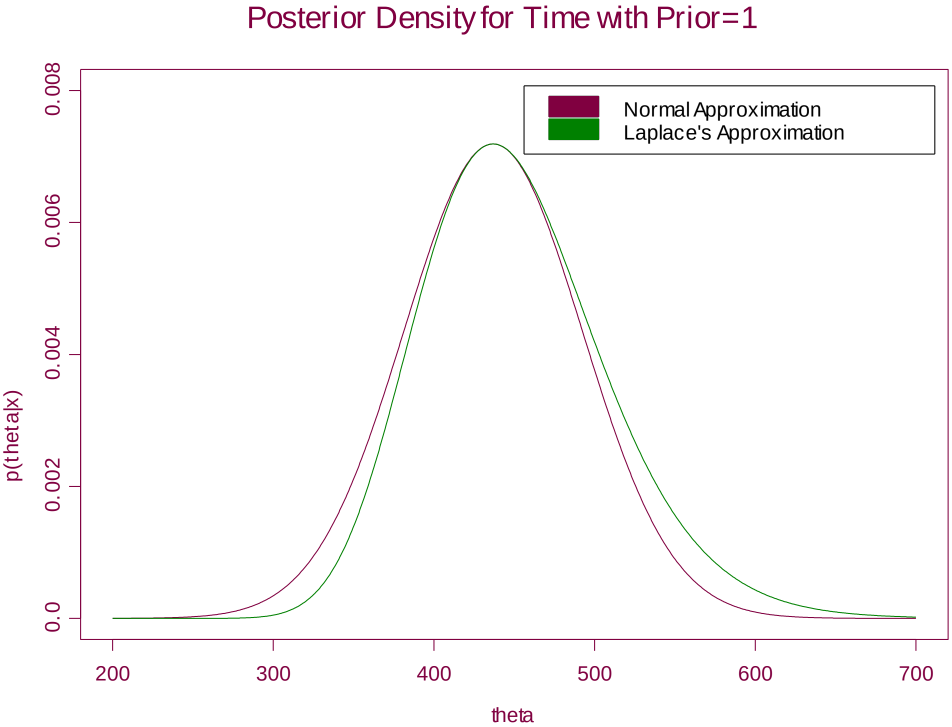


Figure 2.3:Comparison between Normal and Laplace's Approximation

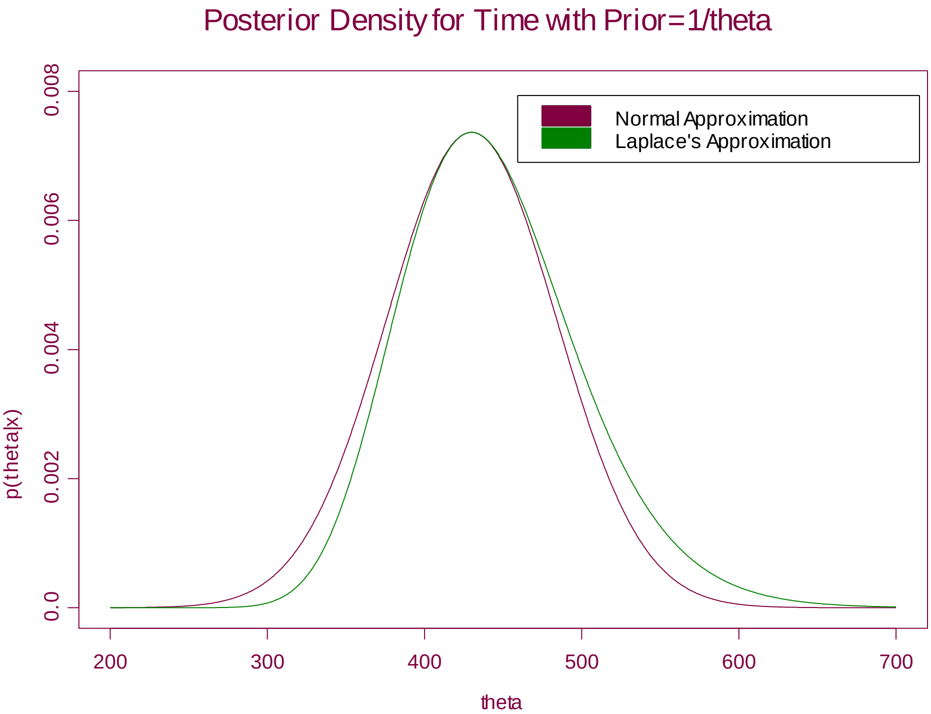


Figure 2.4: Comparison between Normal and Laplace,s Approximation

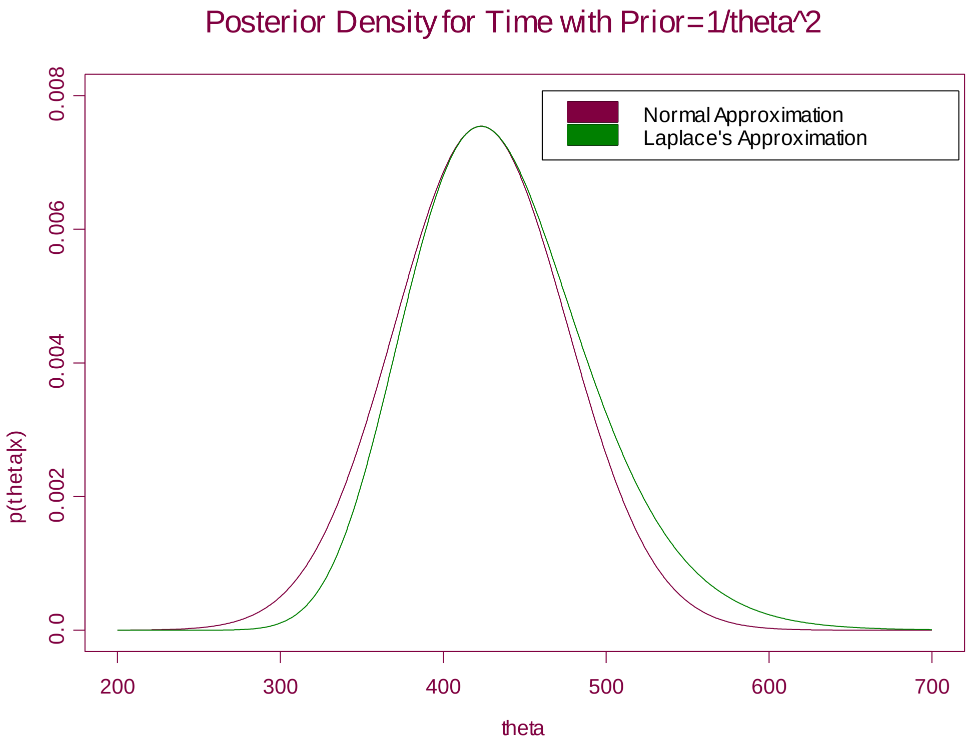


Figure 2.5: Comparison between Normal and Laplace's Approximation

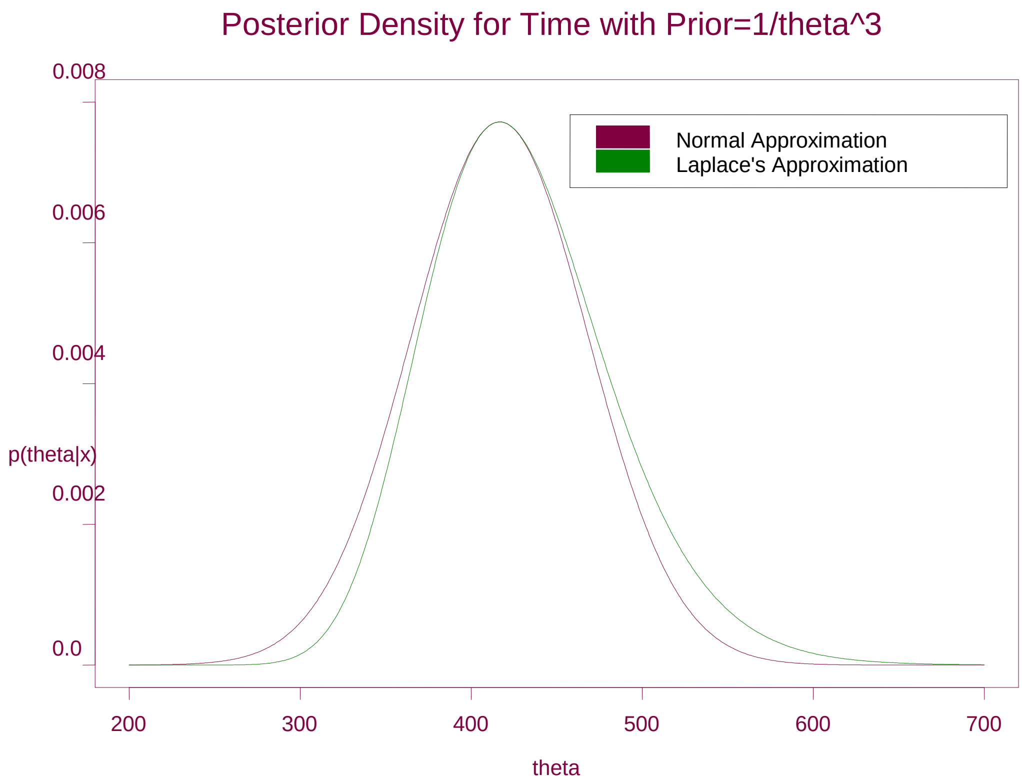


Figure: 2.6: Comparison between Normal and Laplace's Approximation

Example 2.2 (Grubbs, F.E., 1971): Nineteen military personnel carriers failed in services for one reason or the other at the following mileages: 162, 200, 271, 302, 393, 508, 539, 629, 706, 777, 884, 1008, 1101, 1182, 1463, 1603, 1984, 2355 and 2880 miles. Numerical and graphical illustrations are implemented in S-PLUS Software for two parameter exponential distribution. Posterior estimates of μ and θ are given in Table 2.3. The graphical representation for marginal posterior densities of μ and θ are shown in Figures 2.7 and 2.8 respectively. Moreover, we have developed the function for estimating parameters μ and θ of two parameter exponential distribution under different priors. Also, functions for graphical representation of the marginal densities of μ and θ under different priors were also developed in S-PLUS.

The posteriors of μ and θ are plotted in figures 2.7 & 2.8 respectively. The posteriors μ are quite robust for varying c in the prior $p(\mu,\theta) \propto \left(\frac{1}{\theta^c}\right)$ while the posteriors of θ are less robust.

Program for estimating parameters μ and θ of two parameter exponential distribution in S-PLUS.

```
Mu.theta<-function(x)
  {
  n<-length(x)
  C<-c(0,1,2)
  x1<-min(x)
  s<-sum(x-x1)
  estimate1<-x1-(s/(n*(n+C-3)))
  estimate2<-s/(n+C-3)
  list(mu=estimate1,theta=estimate2)
   }
> x<-c(162, 200, 271, 302, 393, 508, 539, 629, 706, 777,
884, 1008, 1101, 1182, 1463, 1603, 1984, 2355,2880 )
>  Mu.theta(x)
```

Table 2.3: Posterior estimates of μ and θ under different priors using S-PLUS.

Prior	Posterior mean of μ	Posterior mean of θ
1	109.7993	991.8125
$\frac{1}{\theta}$	112.8700	933.4706
$\frac{1}{\theta^2}$	115.5994	881.6111

Function for graphical representation of the marginal density of μ under different priors in S-PLUS.

```
mu.plot<-function(x)
{
n<-length(x)
x1<-min(x)
s<-sum(x-x1)
mu<-seq(0,160)
pmu<-(n*(n-2))*(s^(n-2))/((s+n*(x1-mu))^(n-1))
plot(mu,pmu,xlab="mu",ylab="p(mu|x)",ylim=c(0,0.022),
main= "Posterior density of mu under different priors",
sub="Figure: 2.7",type="l",lty=1,col=2)
pmu1<-(n*(n-1))*(s^(n-1))/((s+n*(x1-mu))^(n))
lines(mu,pmu1,lty=2,col=3)
pmu2<-(n*(n))*(s^(n))/((s+n*(x1-mu))^(n+1))
lines(mu,pmu2,lty=3,col=4)
}
> x<-c(162, 200, 271, 302, 393, 508, 539, 629, 706, 777,
884, 1008, 1101, 1182, 1463, 1603, 1984, 2355,2880 )
> Mu.plot(x)
> leg.names<-c("Prior=1","Prior=1/theta","Prior=1/theta^2")
> legend(locator(1),leg.names,col=2:4)
```

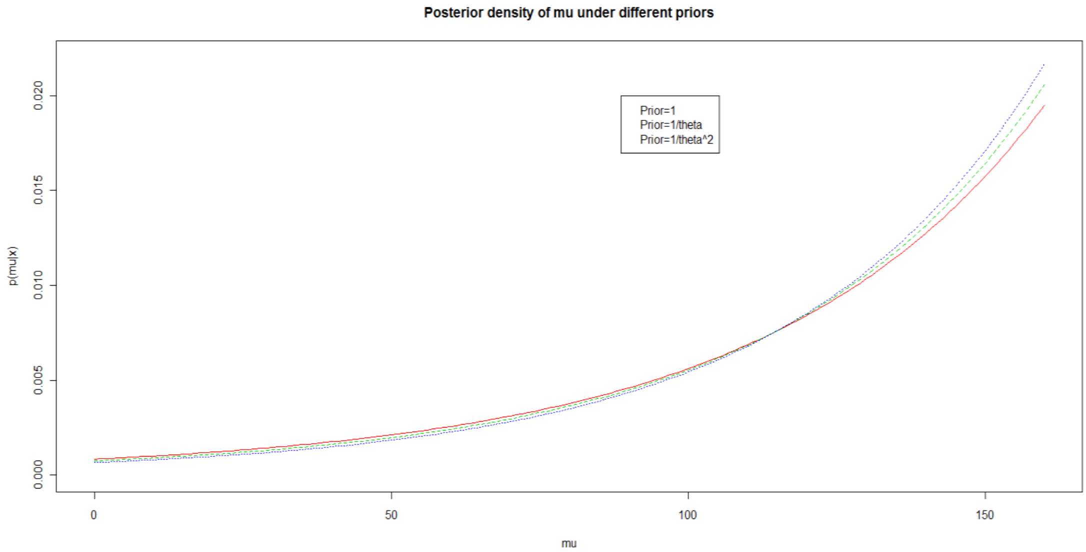


Figure: 2.7

Function for graphical representation of the marginal density of θ under different priors.

```
theta.plot<-function(x)
{
n<-length(x)
x1<-min(x)
s<-sum(x-x1)
theta<-seq(50,1700)
ptheta<- (s^(n-2))*(exp(-s/theta))/((gamma(n-
2))*(theta^(n-1)))
ptheta1<- (s^(n-1))*(exp(-s/theta))/((gamma(n-
1))*(theta^(n)))
ptheta2<- (s^n)*(exp(-
s/theta))/((gamma(n))*(theta^(n+1)))
plot(theta,ptheta,xlab="theta",ylab="p(theta|y)",ylim=c(0
,.0022),
main= "Posterior density of theta under different
priors",sub="Figure: 2.8",type="l",lty=1, col=2)
lines(theta,ptheta1,lty=2,col=3)
lines(theta,ptheta2,lty=3, col=4)
}
>x<-c(162, 200, 271, 302, 393, 508, 539, 629, 706, 777,
884,1008, 1101,1182, 1463, 1603, 1984, 2355,2880 )
```

```
>theta.plot(x)
 >leg.names<-
c("Prior=1","Prior=1/theta","Prior=1/theta^2")
 legend(locator(1),leg.names,col=2:4)
```

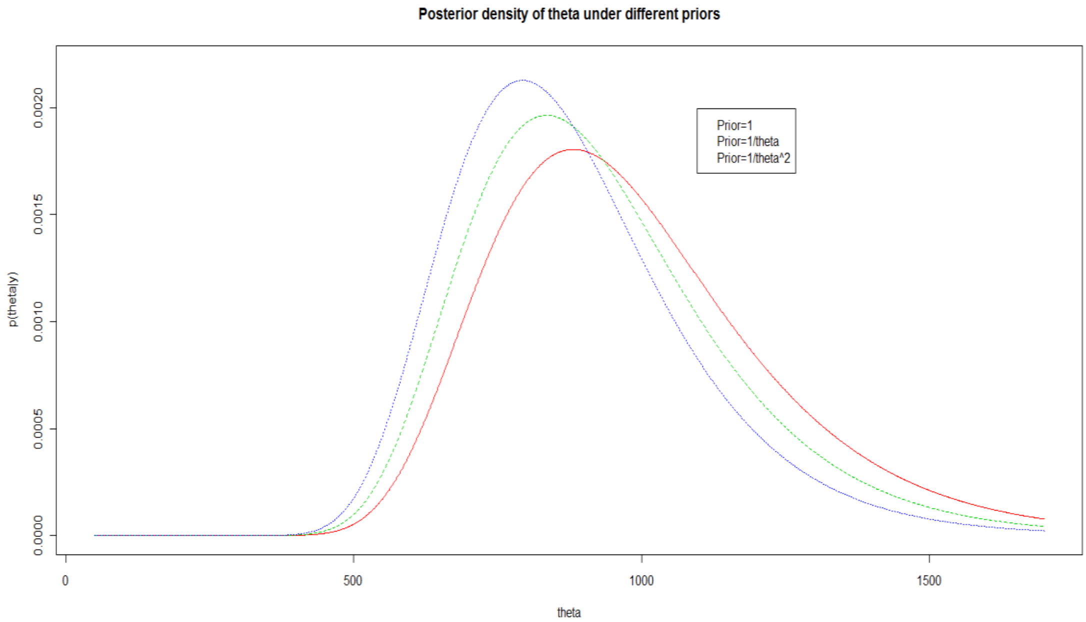


Figure: 2.8

CHAPTER – 3

POSTERIOR APPROXIMATIONS TO GAMMA DISTRIBUTION

3.1 Introduction:

Gamma distribution has been quite extensively used as a lifetime model, though not nearly as much as the weibull distribution. The gamma distribution is most widely used model for precipitation data. It does fit a wide variety of lifetime data adequately, besides failure process models that lead to it. It also arises in some situations involving the exponential distribution; because of the well known result that sums of independent and identically distributed (iid) exponential random variables have a gamma distribution. Inference for gamma model has been considered by Engelhard and Bain (1978), choa and Glaser (1978) and others for complete data case. Prentice (2002); Lawless (2003); Zaman et al. (2005); Jamali et al. (2006); Saal et al. (2008); S.P.Ahmad (2006) & Ahmad et al. (2011) has made significant contributions.

The gamma distribution has pdf of the form

$$f(x;\alpha,\beta)=\frac{1}{\Gamma(\beta)}\frac{x^{\beta-1}}{\alpha^{\beta}}\exp\left\{\frac{-x}{\alpha}\right\};x>0;\ \alpha,\beta>0 \qquad (3.1.1)$$

where $\alpha,\beta>0$ are the parameter, α is a scale parameter and β is sometimes called the index or shape parameter. $\Gamma(\beta)$ Is the well known gamma function which for integral values of equals $(\beta-1)$. The gamma distribution with $\beta=1$ reduces to the one parameter exponential distribution & has pdf of the form

$$f(x)=\frac{1}{\alpha}\exp\left\{\frac{-x}{\alpha}\right\};x>0;\ \alpha>0$$

The gamma distribution with $\alpha=1$ is called the one parameter gamma distribution and has pdf

$$f(x;\beta)=\frac{x^{\beta-1}}{\Gamma(\beta)}\exp\{-x\};x>0;\ \beta>0$$

The moments of a r.v X following gamma distribution can be obtained as

$E(X)=\alpha\beta$ and $V(X)=\alpha\beta^2$

The hazard function of Gamma can be increasing, decreasing or constant depending on $\alpha > 1$, $\alpha < 1$ or $\alpha = 1$ respectively. The exponential distribution the hazard rate is constant ($1/\alpha$) and, therefore, the gamma distribution immediately provide a generalization of the exponential distribution in their distribution.

For the integral value of β, gamma distribution arises as a sum of β independent identically distributed exponential random variables. Therefore, if β items were test and it was assumed that the failure time distribution in exponential with parameter α, then the total time on test(or total of life times) would be a gamma variable with parameter β and α. In failure censored case, with the experiment terminating at the ($\beta - \alpha$) the failure, the total time on test would be distributed as gamma with parameters ($\beta - \alpha$) & α.

3.2 Estimation of parameters of gamma distribution with complete sample:

Let $x_1, x_2, \ldots\ldots, x_n$ be an iid samples from gamma distribution (3.1.1), and the likelihood function is given by

$$L(x \mid \alpha, \beta) = \frac{1}{(\Gamma(\beta))^n} \frac{1}{(\alpha)^{n\beta}} \exp\left\{\frac{-\sum_{i=1}^{n} x_i}{\alpha}\right\}\left(\prod_{i=1}^{n} x_i^{\beta-1}\right) \qquad (3.2.1)$$

Case I: when β is known:

We will first consider the case when β is known and the only unknown parameter is α.

Taking log on both sides of equation (3.2.1) we get

$$\log L(x \mid \alpha, \beta) = -n \log \Gamma(\beta) - n\beta \log \alpha - \frac{1}{\alpha}\sum_{i=1}^{n} x_i + (\beta - 1)\sum_{i=1}^{n} \log x_i$$

$$\therefore \quad \frac{\partial}{\partial \alpha} \log L(x \mid \alpha, \beta) = 0$$

$$\Rightarrow \qquad \frac{-n\beta}{\alpha}+\frac{1}{\alpha^2}\sum_{i=1}^{n}x_i=0$$

(3.2.2)

$$\Rightarrow \quad \hat{\alpha}=\frac{\bar{x}}{\beta}$$

We can easily verify that $E(\hat{\alpha})=E\left(\frac{\sum_{i=1}^{n}x_i}{n\beta}\right)=\alpha$

and $\operatorname{var}(\hat{\alpha})=V\left(\frac{1}{n\beta}\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2\beta^2}\sum_{i=1}^{n}V(x_i)=\frac{\alpha^2}{n\beta}$

The distribution of $\hat{\alpha}$ in gamma with parameters $n\beta$ & $\frac{\alpha}{n\beta}$. This follows from the fact that $x_1, x_2, \ldots\ldots, x_n$ are iid gamma random variables with parameters β, α and therefore $\sum_{i=1}^{n}x_i$ is gamma with parameters $n\beta$ & α , & $\hat{\alpha}=\frac{\sum_{i=1}^{n}x_i}{n\beta}$ is gamma with parameters $n\beta$ & $\frac{\alpha}{n\beta}$.

Thus the pdf of $\hat{\alpha}$ is given by

$$f(\hat{\alpha}\mid\beta,x)=\frac{1}{\Gamma(n\beta)}\frac{1}{(\alpha/n\beta)^{n\beta}}\exp\left\{\frac{\hat{\alpha}}{\alpha/n\beta}\right\}(\hat{\alpha})^{n\beta-1}$$

(3.2.3)

From (3.2.1) and (3.2.3) it immediately follows

$$\frac{L(x_1,x_2,\ldots\ldots,x_n\mid\beta,\alpha)}{g(\hat{\alpha}\mid\beta,\alpha)}=\frac{\Gamma(n\beta)}{[\Gamma\beta]^n}\frac{1}{(n\beta)^{n\beta}}\frac{\prod_{i=1}^{n}x_i^{\beta-1}}{(\hat{\alpha})^{n\beta-1}}$$

which is independent of the unknown parameter α. Thus $\hat{\alpha}$ or equivalently the total time $\sum_{i=1}^{n} x_i$ is sufficient for α. In the gamma model we may note that the mean life is $\beta\alpha$ and if we are interested in estimating $\beta\alpha$, then the MLE and UMVUE of are identical and are given by the sample mean $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Case ii:-we consider the case when $\beta \,\&\, \alpha$ are both unknown:

Let $x_1, x_2, \ldots\ldots, x_n$ be an iid sample from gamma distribution (3.1.1), and then likelihood is defined as

$$L(x \mid \alpha, \beta) = \frac{1}{\alpha^{n\beta}\Gamma(\beta)^n}\left(\prod_{i=1}^{n} x_i^{\beta-1}\right)\exp\left(-\frac{\sum_{i=1}^{n} x_i}{\alpha}\right)$$

$$L(x \mid \alpha, \beta) = \frac{1}{\alpha^{n\beta}\Gamma(\beta)^n}(\tilde{x})^{n(\beta-1)}\exp\left(-\frac{n\bar{x}}{\alpha}\right) \tag{3.2.4}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\tilde{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ are arithmetic and geometric means, respectively.

The log-likelihood is given by

$$\log L(x \mid \alpha, \beta) = -n\beta \log \alpha - n \log \Gamma(\beta) + n(\beta - 1)\log \tilde{x} - \frac{n\bar{x}}{\alpha} \tag{3.2.5}$$

Differentiating (3.2.5) with respect to α we have

$$\frac{\partial \log L(x \mid \alpha, \beta)}{\partial \alpha} = \frac{-n\beta}{\alpha} + \frac{n\bar{x}}{\alpha^2}$$

$$\Rightarrow \alpha\beta = \bar{x} \tag{3.2.6}$$

Equation (3.2.5) can be written as

$$\log L(x \mid \alpha, \beta) = -n\beta \log \alpha + n\beta \log \tilde{x} - n \log \Gamma(\beta) + n \log \tilde{x} - \frac{n\bar{x}}{\alpha}$$

$$= -n\beta \log\left(\frac{\alpha}{\tilde{x}}\right) - n \log \Gamma(\beta) + n \log \tilde{x} - \frac{n\bar{x}}{\alpha} \qquad (3.2.7)$$

Differentiating equation (3.2.5) w.r.t β and equating to zero we have

$$\frac{\partial \log L(x \mid \alpha, \beta)}{\partial \beta} = -n \log\left(\frac{\alpha}{\tilde{x}}\right) - n \frac{\partial}{\partial \beta} \log \Gamma(\beta)$$

$$= -\left[\log\left(\frac{\bar{x}}{\beta}\right) - \log \tilde{x}\right] - \psi(\beta) = 0$$

Where $\psi(\beta) = \frac{\partial \log \Gamma(\beta)}{\partial \beta} = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$ is a dia-gamma function.

$$\Rightarrow -[\log \bar{x} - \log \tilde{x} - \log \beta] - \psi(\beta) = 0$$

$$\Rightarrow -\log\left(\frac{\bar{x}}{\tilde{x}}\right) + \log \beta - \psi(\beta) = 0$$

$$\Rightarrow \log \beta - \psi(\beta) = \log\left(\frac{\bar{x}}{\tilde{x}}\right) \qquad (3.2.8)$$

$\psi'(\beta) = \frac{\partial \psi(\beta)}{\partial \beta}$ is termed as tri-gamma function. These functions can be approximated well as

$$\psi(\beta) = \log \beta - \frac{1}{2\beta} - \frac{1}{12\beta^2} + \frac{1}{120\beta^3} + \ldots$$

and $$\psi'(\beta) = \frac{1}{\beta} + \frac{1}{2\beta^2} + \frac{1}{6\beta^3} - \frac{1}{30\beta^5} + \ldots$$

These values are required to implement Newton's method of optimization. However, this method is difficult to implement as compared to a very close approximation discussed by Johnson and Kotz (1970). The maximum estimate of β can be approximated as

$$\hat{\beta} = s^{-1}(0.5000876 + 0.1648852s - 0.0544274s^2) \text{ for } 0 < s \le 0.55722 \qquad (3.2.9)$$

&

$$\hat{\beta} = s^{-1}(17.79728 + 11.968477s + s^2)^{-1} \times (8.898919 + 9.05995s + 0.9775373s^2)$$

$$\text{for } 0.557220 < s \le 17$$

(3.2.10)

where $s = \frac{\bar{x}}{\hat{\beta}}$. If the value of *s* ranges from 0 to 0.55722, then value of $\hat{\beta}$ is given by (3.2.9) and if it lies between 0.55722 and 17, then it is given by (3.2.10). Once $\hat{\beta}$ is obtained, we can find $\hat{\alpha}$ from

$$\hat{\alpha} = \frac{\bar{x}}{\hat{\beta}} \quad .$$

For large values of β, we can use the approximation

$$\frac{\partial \log \Gamma(\beta)}{\partial \beta} \approx \log\beta - \frac{1}{2\beta} \quad \text{so that}$$

$$\hat{\beta} = \frac{1}{2(\log\bar{x} - \log\tilde{x})} \quad \text{and} \quad \hat{\alpha} = \frac{\bar{x}}{\hat{\beta}} \, .$$

These estimates are essentially needed for starting iterations. The expression for the variance covariance matrix Σ of these estimates could be obtained by using the asymptotic properties of MLE. Using the general theory of MLE, one can show that asymptotically $\left(\hat{\beta}, \hat{\alpha}\right)$ is distributed as bivariate normal with mean α & β respectively and the variance covariance matrix is given by

$$\sum = -E\begin{bmatrix} \frac{\partial^2 \log L}{\partial \beta^2} & \frac{\partial^2 \log L}{\partial \beta\, \partial \alpha} \\ \frac{\partial^2 \log L}{\partial \alpha\, \partial \beta} & \frac{\partial^2 \log L}{\partial \alpha^2} \end{bmatrix}^{-1}$$

Where $\frac{\partial^2 \log L}{\partial \beta^2} = \frac{-n\,\partial^2}{\partial \beta^2}[\log \Gamma \beta] = -n\Psi'(\beta)$ is called the gamma function.

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \frac{n\beta}{\alpha^2} - \frac{2n\overline{x}}{\alpha^3} = \frac{n\beta}{\alpha^2} - \frac{2n\alpha\beta}{\alpha^3} = \frac{-n\beta}{\alpha^2}$$

$$\frac{\partial^2 \log L}{\partial \beta\, \partial \alpha} = \frac{-n}{\alpha}$$

$$\therefore \sum = -E\begin{bmatrix} -n\Psi'(\beta) & \frac{-n}{\alpha} \\ \frac{-n}{\alpha} & \frac{-n\beta}{\alpha^2} \end{bmatrix}^{-1} = E\begin{bmatrix} n\Psi'(\beta) & \frac{n}{\alpha} \\ \frac{n}{\alpha} & \frac{n\beta}{\alpha^2} \end{bmatrix}^{-1}$$

$$= \frac{1}{n}\begin{bmatrix} \frac{\beta/\alpha^2}{D} & \frac{-1/\alpha}{D} \\ \frac{-1/\alpha}{D} & \frac{\Psi'(\beta)}{D} \end{bmatrix}$$

Where $D = \Psi'(\beta)\frac{\beta}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}[\Psi'(\beta)\beta - 1] = \frac{\Delta\beta}{\alpha^2}$

$$\sum = \frac{1}{n}\begin{bmatrix} \frac{\beta}{\Delta\beta} & \frac{-\alpha}{\Delta\beta} \\ \frac{-\alpha}{\Delta\beta} & \frac{\Psi'(\beta)\alpha^2}{\Delta\beta} \end{bmatrix}$$

$$\therefore V(\hat{\beta}) = \frac{\beta}{n\Delta\beta} \;\; \& \;\; V(\hat{\alpha}) = \frac{\Psi'(\beta)\alpha^2}{n\Delta\beta}$$

If $x_1, x_2, \ldots\ldots, x_n$ are independently and identically distributed as (3.1.1), it follows that

$$\frac{\sum_{i=1}^{n} X_i}{\alpha} \sim \gamma(n\beta)$$

or $\frac{n\overline{X}}{\alpha} \sim \gamma(n\beta)$

or $\dfrac{\overline{X}}{\alpha/n} \sim \gamma(n\beta)$

and hence $\overline{X} \sim \gamma\left(n\beta, \dfrac{\alpha}{n}\right)$

Since $\dfrac{n\overline{X}}{\alpha} \sim \gamma(n\beta)$

$$E\left(\frac{n\overline{X}}{\alpha}\right) = Var\left(\frac{n\overline{X}}{\alpha}\right) = (n\beta)$$

From which it follows that $E(\overline{X}) = \alpha\beta,\ Var(\overline{X}) = \left(\dfrac{\beta\alpha^2}{n}\right)$

For large sample $\overline{X}$ is asymptotically normal with that same mean and variance $\alpha\beta,\ \&\left(\dfrac{\beta\alpha^2}{n}\right)$ respectively.

3.3 Approximation of Gamma Distribution Based on Posterior Modes:

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. We make use of Normal and Laplace's methods of approximation as discussed by Rubin and Schenker (1987) and Tierney and Kadane (1986).

From (3.2.5) the log-likelihood is defined as:

$$\log L(x \mid \alpha, \beta) = -n\beta \log\alpha - n\log\Gamma(\beta) + n(\beta - 1)\log\tilde{x} - \frac{n\overline{x}}{\alpha}$$

We take partial derivatives with respect to α and β .

$$l_\alpha = \frac{\partial \log L(x \mid \alpha, \beta)}{\partial\alpha} = \frac{-n\beta}{\alpha} + \frac{n\overline{x}}{\alpha^2}$$

$$l_{\beta} = \frac{\partial \log L(x \mid \alpha, \beta)}{\partial \beta} = n \log\left(\frac{\bar{x}}{\alpha\beta}\right) + \frac{n}{\beta}$$

$$l_{\alpha\beta} = \frac{\partial^2 \log L(x \mid \alpha, \beta)}{\partial \alpha \, \partial \beta} = -\frac{n}{\alpha}$$

$$l_{\beta\alpha} = \frac{\partial^2 \log L(x \mid \alpha, \beta)}{\partial \beta \, \partial \alpha} = -\frac{n}{\alpha}$$

$$l_{\alpha\alpha} = \frac{\partial^2 \log L(x \mid \alpha, \beta)}{\partial \alpha^2} = \frac{n\beta}{\alpha^2} - \frac{2n\bar{x}}{\alpha^3}$$

$$l_{\beta\beta} = \frac{\partial^2 \log L(x \mid \alpha, \beta)}{\partial \beta^2} = -\frac{n(\beta - 1)}{\beta^2}$$

We follow the standard approach of Box and Tiao (1973), Gelman et al. (1995), we assume that a priori α and k are approximately independent, so that $g(\alpha, \beta) \cong g(\alpha) g(\beta)$. Where $g(\alpha)$ and $g(\beta)$ are priors for α and β. Using Bayes theorem, the posterior density $P(\alpha, \beta \mid x)$ is given by

$$P(\alpha, \beta \mid x) \propto \prod_{i=1}^{n} f(x_i \mid \alpha, \beta)\, g(\alpha)\, g(\beta) \qquad (3.3.1)$$

The log-posterior is given by

$$\log P(\alpha, \beta \mid x) = \log \prod_{i=1}^{n} f(x_i \mid \alpha, \beta) + \log g(\alpha) + \log g(\beta)$$

or $\quad \mathrm{l}^*(\alpha, \beta) = \mathrm{l}(\alpha, \beta) + \log \mathrm{g}(\alpha) + \log \mathrm{g}(\beta) \qquad (3.3.2)$

For a prior $\mathrm{g}(\alpha, \beta) \cong \mathrm{g}(\alpha)\, \mathrm{g}(\beta) = 1$, we have

$$\mathrm{l}^*_{\alpha} = \mathrm{l}_{\alpha}\ , \quad \mathrm{l}^*_{\beta} = \mathrm{l}_{\beta}, \quad \mathrm{l}^*_{\alpha\beta} = \mathrm{l}_{\alpha\beta}\ , \quad \mathrm{l}^*_{\beta\alpha} = \mathrm{l}_{\beta\alpha}, \quad \mathrm{l}^*_{\alpha\alpha} = \mathrm{l}_{\alpha\alpha} \text{ and } \mathrm{l}^*_{\beta\beta} = \mathrm{l}_{\beta\beta}.$$

The posterior mode is obtained by maximizing (3.3.2) with respect to α and β. The score vector of log posterior is given by

$$\mathrm{U}(\alpha, \beta) = (\mathrm{l}^*_{\alpha}, \mathrm{l}^*_{\beta})^{\mathrm{T}}$$

and Hessian matrix of log posterior is

$$H(\alpha,\beta)=\begin{bmatrix} l^*_{\alpha\alpha} & l^*_{\alpha\beta} \\ l^*_{\beta\alpha} & l^*_{\beta\beta} \end{bmatrix}$$

Posterior mode $(\hat{\alpha},\hat{\beta})$ can be obtained from Newton-Raphson iteration scheme

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}=\begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix}-H^{-1}(\alpha_0,\beta_0)\begin{bmatrix} l^*_{\alpha} \\ l^*_{\beta} \end{bmatrix} \qquad (3.3.3)$$

Consequently, modal variance Σ can be obtained as

$$I^{-1}(\hat{\alpha},\hat{\beta})=-H^{-1}(\hat{\alpha},\hat{\beta})\,.$$

$P(\alpha,\beta\,|\,x)$ can be used for drawing inference about α and β simultaneously.

Using normal approximation, we can write directly a bivariate normal approximation of (3.3.1) as

$$P(\alpha,\beta\,|\,x)\cong N_2((\hat{\alpha},\hat{\beta})^T, I^{-1}(\hat{\alpha},\hat{\beta}))$$

Similarly, we can write Bayesian analog of likelihood ratio criterion as

$$-2[l^*(\alpha,\beta)-l^*(\hat{\alpha},\hat{\beta})]\sim\chi^2_2$$

Using Laplace's approximation, we can write (3.3.1) as

$$P(\alpha,\beta\,|\,x)\cong(2\pi)^{-1}\,|\,I(\hat{\alpha},\hat{\beta})\,|^{\frac{1}{2}}\,\exp[l^*(\alpha,\beta)-l^*(\hat{\alpha},\hat{\beta})]$$

The marginal Bayesian inference about α and β is to be based on marginal posterior densities of these parameters. Marginal posterior for α can be obtained after integrating out $P(\alpha,\beta\,|\,x)$ with respect to β, i.e. $P(\alpha\,|\,x)=\int_0^\infty P(\alpha,\beta\,|\,x)\,d\beta$.

Similarly, marginal posterior of β can be obtained as

$$P(\beta \mid x) = \int_0^\infty p(\alpha, \beta \mid x)\, d\alpha$$

We can write normal approximation of marginal posterior $p(\alpha \mid x)$ as

$$P(\alpha \mid x) = N_1(\hat{\alpha}, I_{11}^{-1})$$

Bayesian analog of likelihood ratio criterion can also be defined as a test criterion as

$$(\alpha - \hat{\alpha})^T I_{11} (\alpha - \hat{\alpha})! \sim \chi_1^2$$

Laplace's approximation of marginal posterior density $p(\alpha \mid x)$ can be given by

$$P(\alpha \mid x) \cong \left[\frac{|I(\hat{\alpha}, \hat{\beta})|}{2\pi\, |I(\alpha, \hat{\beta}(\alpha))|} \right]^{\frac{1}{2}} \exp[l^*(\alpha, \hat{\beta}(\alpha)) - l^*(\hat{\alpha}, \hat{\beta})]$$

Similarly, $P(\beta \mid x)$ can be approximated and results corresponding to normal and Laplace's approximation can be written as

$$P(\beta \mid x) = N_1(\hat{\beta}, I_{22}^{-1})$$

or equivalently,

$$(\beta - \hat{\beta})^T I_{22} (\beta - \hat{\beta}) \sim \chi_1^2$$

$$P(\beta \mid x) \cong \left[\frac{|I(\hat{\alpha}, \hat{\beta})|}{2\pi\, |I(\hat{\alpha}(\beta), \beta)|} \right]^{\frac{1}{2}} \exp[l^*(\hat{\alpha}(\beta), \beta) - l^*(\hat{\alpha}, \hat{\beta})]$$

The results can be seen in Ahmad(2006) & Ahmad et al. (2011)

Example 3.1: The numerical and graphical illustration of posterior densities of the parameters of interest conveys a very convincing and comprehensive picture of Bayesian data analysis. We have developed several programs using S-PLUS and R softwares for gamma distribution. These programmes illustrate the strength of Bayesian methods in various practical situations. Soil samples were collected from rice growing areas as well as from orchards of Kashmir valley and were analyzed from some relevant parameters. Ahmad et.al, 2011 studied available Potassium in the soil of Kashmir valley. The posterior mode and standard errors of parameters α and β of gamma distribution are presented in Table 3.1. Graphical display of posterior for α and β using Normal approximation are shown in Figures 3.1 to 3.6, whereas Laplace's approximation for α and β are shown in Figures 3.7 to 3.12. Figures 3.13 to 3.15 and 3.16 to 3.18 contains Normal approximation of posterior in addition to Laplace's approximation for parameters α and k respectively. The graph shows that the two approximations are in close agreement.

 # Bayesian Analysis of Gamma Distribution with different Priors in SPLUS.

 # Prior=1.

```
library(Mass,first=T)
n<-length(x)

ngam1<-deriv3(~-log(x^(k1))+y/alpha+b*log(alpha)+lgamma(b),

              c("alpha","b"),function(x,alpha,b)NULL)

y<-dbmdata$Potassium

y<-as.vector(x)

fitgam1<-ms(~ngam1(x,alpha,b),start=c(alpha=66,b=12),data=dbmdata)

post.std<-sqrt(diag(summary(fitgam1)$Information))

summary(fitgam1)

post.std

> summary(fitgam1)

Final value: 8382.159

Solution:

           Par.          Grad. Hessian.alph  Hessian.b

alpha  60.055763 -1.113207e-013     1.197854  23.428226

    b   3.070571 -2.176259e-011    23.428226 540.775513

Information:

          alph          b
```

```
alph  5.468549 -0.2369160

   b -0.236916  0.0121132

Convergence: RELATIVE FUNCTION CONVERGENCE.

Computations done:

  Iterations Function Gradient

      8       10        9

> post.std

[1] 2.338493 0.110060
```

 # Prior=1/b.

```
library(Mass,first=T)

n<-length(x)

ngam1<-deriv3(~-log(x^(k-1))+x/alpha+b*log(alpha)+lgamma(b)-
log(1/b)/n,c("alpha","b"),function(x,alpha,b)NULL)

y<-dbmdata$Potassium

y<-as.vector(x)

fitgam1<-ms(~ngam1(x,alpha,b),start=c(alpha=66,b=12),data=dbmdata)

post.std<-sqrt(diag(summary(fitgam1)$Information))

summary(fitgam1)

post.std

> summary(fitgam1)

Final value: 8383.281

Solution:

           Par.     Grad. Hessian.alph  Hessian.b

alpha 60.133015  4.540118e-014     1.193243  23.398128

    b  3.066626 -8.013770e-012    23.398128 541.480975

Information:
```

```
          alph          b
alph  5.4891249 -0.2371925
   b -0.2371925  0.0120962
Convergence: RELATIVE FUNCTION CONVERGENCE.
Computations done:
 Iterations Function Gradient
     8       10        9
> post.std
[1] 2.3428882 0.1099827
```

 # Prior=1/(alpha*b).

```
library(Mass,first=T)
n<-length(x)
ngam1<-deriv3(~-log(x^(k-1))+x/alpha+b*log(alpha)+lgamma(b)-

log(1/(alpha*b))/n,c("alpha","b"),function(x,alpha,b)NULL)
y<-dbmdata$Potassium
y<-as.vector(x)
fitgam1<-ms(~ngam1(x,alpha,b),start=c(alpha=66,b=12),data=dbmdata)
post.std<-sqrt(diag(summary(fitgam1)$Information))
summary(fitgam1)
post.std
Final value: 8387.376
Solution:
           Par.          Grad. Hessian.alph  Hessian.b
alpha 60.041875 -3.260959e-013     1.198685  23.433645
    b  3.070571 -3.199555e-011    23.433645 540.669544
```

```
Information:

           alph          b

alph  5.4636845 -0.2368065

   b -0.2368065  0.0121132

Convergence: BOTH X- AND RELATIVE FUNCTION CONVERGENCE

Computations done:

 Iterations Function Gradient

      9       11        9

> post.std

[1] 2.337453 0.110060
```

Table 3.1: Posterior mode and Posterior standard error of Gamma distribution with different priors.

Prior	Posterior mode		Posterior Standard error	
	α	β	α	β
1	60.065413	3.070154	2.3420626	0.1101784
1/b	60.14267	3.06621	2.3464633	0.1101011
1/b*alpha	60.051520	3.070153	2.3410191	0.1101783

Normal Approximation to parameters alpha and k of Gamma Distribution using

different priors in S-PLUS and R.

 # Normal Approximation of alpha of Gamma Distribution with different priors in S-PLUS and R.

```
Norm.app<-function(x)

{

n<-length(x)
```

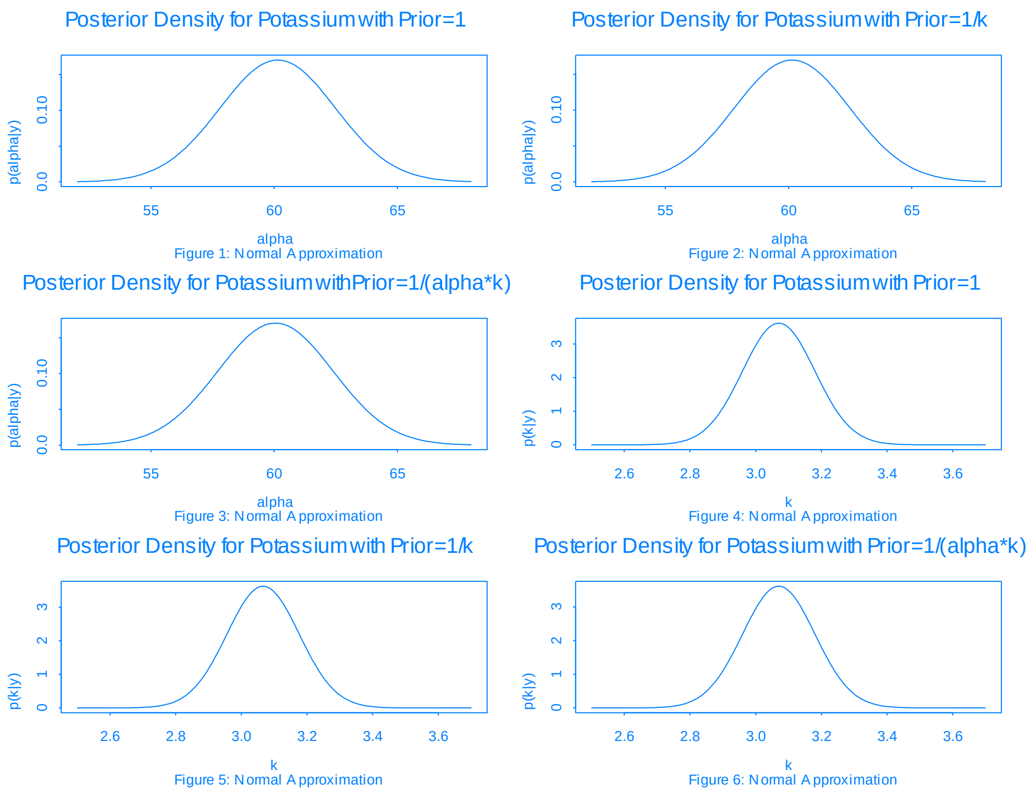
```
alpha<-seq(52,68,length=150)
plot(alpha,dnorm(alpha,mean=60.14267,sd=2.3420626),xlab="alpha",
     ylab="p(alpha|y)",main="Posterior Density for Potassium with
     Prior=1",sub="Figure 3.1: Normal Approximation",type="l",
     col=4)
plot(alpha,dnorm(alpha,mean=60.14267,sd=2.3464633),xlab="alpha",
     ylab="p(alpha|x)",main="Posterior Density for Potassium with
     Prior=1/b",sub="Figure 3.2: Normal Approximation",type="l",
     col=4)
plot(alpha,dnorm(alpha,mean=60.051520,sd=2.3410191),xlab="alpha",
     ylab="p(alpha|x)",main="Posterior Density for Potassium with
     Prior=1/(alpha*b)",sub="Figure 3.3: Normal Approximation",
     type="l",col=4)
}
Norm.app(dbmdata$Potassium)
```

 # Normal Approximation of k of Gamma Distribution with different priors.

 # S-PLUS and R.

```
Norm.app<-function(x)
{
n<-length(x)
k<-seq(2.5,3.7,length=150)
plot(b,dnorm(b,mean=3.070154,sd=0.1101784),xlab="b",ylab="p(b|x)",
     main="Posterior Density for Potassium with Prior=1",
     sub="Figure 3.4: Normal Approximation",type="l",col=4)
plot(x,dnorm(b,mean=3.06621,sd=0.1101011),xlab="b",ylab="p(b|x)",
     main="Posterior Density for Potassium with Prior=1/b",
```

```
     sub="Figure 3.5: Normal Approximation",type="l",col=4)

plot(b,dnorm(b,mean=3.070153,sd=0.1101783),xlab="b",ylab="p(b|x)",

     main="Posterior Density for Potassium with Prior=1/(alpha*b)",

     sub="Figure 3.6: Normal Approximation",type="l",col=4)

}

Norm.app(dbmdata$Potassium)
```



Laplace's Approximation to parameters alpha and k of Gamma Distribution using different priors in S-PLUS and R.

 # Laplace's Approximation of alpha of Gamma Distribution with different priors.

 # S-PLUS and R.

```
Lap.app<-function(x)

 {

 alpha<-seq(52,68,length=9)

 dk<-c(468.8552,486.7220,504.5719,522.4035,540.2163,558.0095,
```

```
      575.7837,593.5376,611.2709)
Lest<-8382.265
Lestb<-c(8389.484,8386.146,8383.925,8382.674,8382.265,8382.593,
         8383.563,8385.097,8387.125)
palpha<-1/sqrt(2*pi)*sqrt( 98.59129/db)*exp(-(Lestb-Lest))
plot(spline(alpha,palpha,n=5*length(alpha),xmin=min(alpha),
     xmax=max(alpha)),xlab="alpha",ylab="p(alpha|x)",
     main="Posterior Density for Potassium with Prior=1",
     sub="Figure 3.7: Laplace's Approximation",type="l",col=4)
db1<-c(468.8675,486.7355,504.5867,522.4197,540.2338,558.0289,
       575.8045,593.5600,611.2952)
Lest1<-8383.386
Lestb1<-c(8390.729,8387.359,8385.107,8383.825,8383.388,8383.687,
          8384.631,8386.140,8388.142)
palpha1<-1/sqrt(2*pi)*sqrt( 98.35003/db1)*exp(-(Lestb1-Lest1))
plot(spline(alpha,palpha1,n=5*length(alpha),xmin=min(alpha),
    xmax=max(alpha)),xlab="alpha",ylab="p(alpha|x)",
    main="Posterior Density for Potassium with Prior=1/b",
    sub="Figure 3.8: Laplace's Approximation",type="l",col=4)
dk2<-c(468.8675,486.7354,504.5866,522.4196,540.2339,558.0290,
       575.8044,593.5601,611.2952)
Lest2<-8387.482
Lestb2<-c(8394.680,8391.348,8389.132,8387.886,8387.482,8387.815,
          8388.790,8390.329,8392.362)
palpha2<-1/sqrt(2*pi)*sqrt( 98.66004/db2)*exp(-(Lestk2-Lest2))
plot(spline(alpha,palpha2,n=5*length(alpha),xmin=min(alpha),
```

```
      xmax=max(alpha)),xlab="alpha",ylab="p(alpha|x)",
main="Posterior Density for Potassium with Prior=1/(alpha*b)",
      sub="Figure 3.9: Laplace's Approximation",type="l",col=4)
 }
Lap.app(dbmdata$Potassium)
```

 # Laplace's Approximation of k of Gamma Distribution with different priors.

 # S-PLUS and R.

```
Lap.app<-function(x)
 {
 b<-seq(2.5,3.7,length=13)
 dalpha<-c(0.646208,0.726896,0.814036,0.907876,1.008664,1.116648,
          1.232073,1.355196,1.486258,1.625509,1.773195,1.929566,
          2.094865)
 Lest<-8382.265
 Lestalpha<-c(8389.725,8392.500,8388.448,8385.477,8383.477,8382.471,
             8382.301,8382.941,8384.339,8386.446,8389.222,8392.626,
             8396.622)
 pk<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha)*exp(-(Lestalpha-Lest))
 plot(spline(b,pb,n=5*length(b),xmin=min(b),xmax=max(b)),
      xlab="b",ylab="p(b|x)",main="Posterior Density for Potassium
   with Prior=1",sub="Figure3.10:Laplace's Approximation",
   type="l", col=4)
dalpha1<-c(0.646208,0.726896,0.814036,0.907876,1.008664,1.116648,
          1.232073,1.355196,1.486258,1.625509,1.773195,1.929566,
          2.094865)
 Lest1<-8383.386
```

```
 Lestalpha1<-
c(8398.641,8393.456,8389.441,8386.507,8384.573,8383.570,


8383.433,8384.104,8385.532,8387.670,8390.475,8393.907,

               8397.930)

 pb1<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha1)*exp(-(Lestalpha1-Lest1))

 plot(spline(b,pb1,n=5*length(b),xmin=min(b),xmax=max(b)),

      xlab="b",ylab="p(b|x)",main="Posterior Density for Potassium
   with Prior=1/b",sub="Figure 3.11: Laplace's Approximation",
   type="l",col=4)

  dalpha2<-c(0.646760,0.727492,0.844679,0.908567,1.009405,1.117441,

            1.232920,1.356099,1.487218,1.626528,1.774276,1.930709,

            2.096072)

 Lest2<-8387.482

 Lestalpha2<-
c(8402.942,8397.717,8393.665,8390.694,8388.726,8387.689,

            8387.519,8388.158,8389.556,8391.663,8394.439,8397.843,

            8401.840)

 pb2<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha2)*exp(-(Lestalpha2-Lest2))

 plot(spline(b,pb2,n=5*length(b),xmin=min(b),xmax=max(b)),

      xlab="b",ylab="p(b|x)",main="Posterior Density for Potassium
   with Prior=1/(alpha*b)",sub="Figure 3.12: Laplace's
   Approximation", type="l",col=4)

}

Lap.app(dbmdata$Potassium)
```

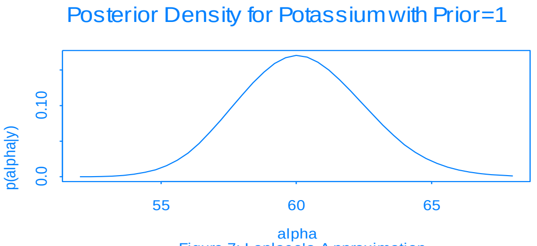


Figure 7: Laplace's Approximation

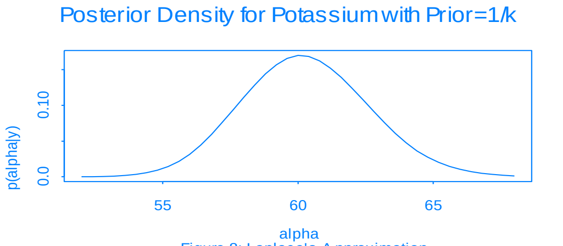


Figure 8: Laplace's Approximation

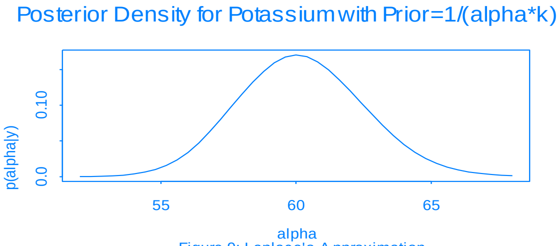


Figure 9: Laplace's Approximation

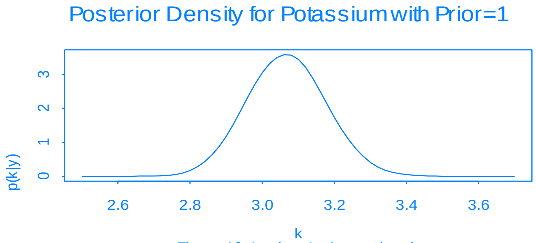


Figure 10: Laplace's Approximation

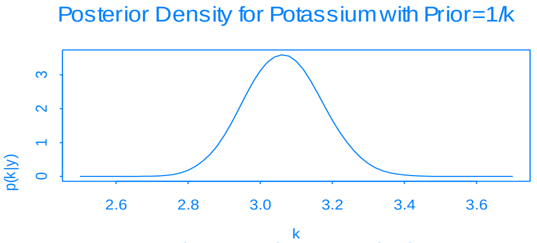


Figure 11: Laplace's Approximation

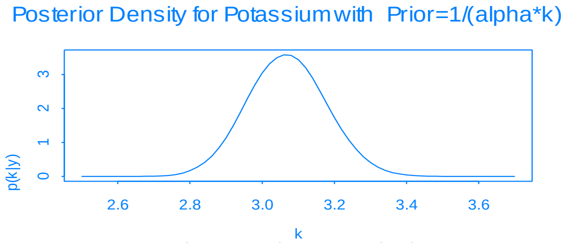


Figure 12: Laplace's Approximation

Comparing Normal and Laplace's Approximation of alpha of Gamma Distribution with different

 # Comparing Normal and Laplace's Approximation of alpha of Gamma distribution with different priors.

 # S-PLUS and R.

```
Norm.Lap<-function(x)

{

n<-length(x)

alpha<-seq(52,68,length=9)

plot(spline(alpha,dnorm(alpha,mean=60.14267,sd=2.3420626 ),


   n=5*length(alpha),xmin=min(alpha),xmax=max(alpha)),xlab="alpha",

      ylab="p(alpha|x)",main="Posterior Density for Potassium with

      Prior=1",sub="Figure 3.13: Comparison between Normal and

      Laplace's Approximation",type="l",col=3)
```

```
db<-c(468.8552,486.7220,504.5719,522.4035,540.2163,558.0095,
      575.7837,593.5376,611.2709)
Lest<-8382.265
Lestb<-c(8389.484,8386.146,8383.925,8382.674,8382.265,8382.593,
         8383.563,8385.097,8387.125)
palpha<-1/sqrt(2*pi)*sqrt( 98.59129/db)*exp(-(Lestb-Lest))
lines(spline(alpha,palpha,n=5*length(alpha),xmin=min(alpha),
      xmax=max(alpha)),col=4)
plot(spline(alpha,dnorm(alpha,mean=60.14267,sd=2.3464633),

   n=5*length(alpha),xmin=min(alpha),xmax=max(alpha)),xlab="alpha",
      ylab="p(alpha|x)",main="Posterior Density for Potassium with
     Prior=1/b",sub="Figure 3.14: Comparison between Normal and
     Laplace's Approximation",type="l",col=3)
db1<-c(468.8675,486.7355,504.5867,522.4197,540.2338,558.0289,
       575.8045,593.5600,611.2952)
Lest1<-8383.386
Lestb1<-c(8390.729,8387.359,8385.107,8383.825,8383.388,8383.687,
           8384.631,8386.140,8388.142)
palpha1<-1/sqrt(2*pi)*sqrt( 98.35003/dk1)*exp(-(Lestb1-Lest1))
lines(spline(alpha,palpha1,n=5*length(alpha),xmin=min(alpha),
      xmax=max(alpha)),col=4)
plot(spline(alpha,dnorm(alpha,mean=60.051520,sd=2.3410191 ),

   n=5*length(alpha),xmin=min(alpha),xmax=max(alpha)),xlab="alpha",
     ylab="p(alpha|x)",main="Posterior Density for Potassium with
     Prior=1/(alpha*b)",sub="Figure 3.15: Comparison between Normal
```

```
    and Laplace's Approximation",type="l",col=3)
 db2<-c(468.8675,486.7354,504.5866,522.4196,540.2339,558.0290,
        575.8044,593.5601,611.2952)
 Lest2<-8387.482
 Lestb2<-c(8394.680,8391.348,8389.132,8387.886,8387.482,8387.815,
          8388.790,8390.329,8392.362)
 palpha2<-1/sqrt(2*pi)*sqrt( 98.66004/db2)*exp(-(Lestb2-Lest2))
 lines(spline(alpha,palpha2,n=5*length(alpha),xmin=min(alpha),
      xmax=max(alpha)),col=4)
 }
Norm.Lap(dbmdata$Potassium)
leg.names<-c("Normal Approximation","Laplace's Approximation")
legend(locator(1),leg.names,fill=3:4)
```

priors using S-PLUS and R.

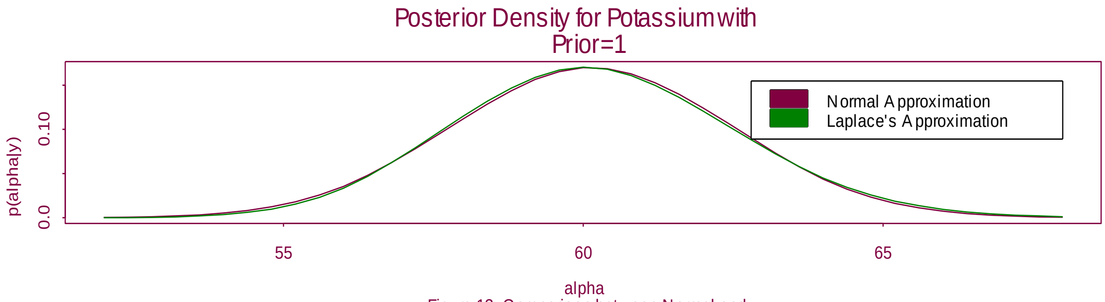


Figure 13: Comparison between Normal and Laplace's Approximation

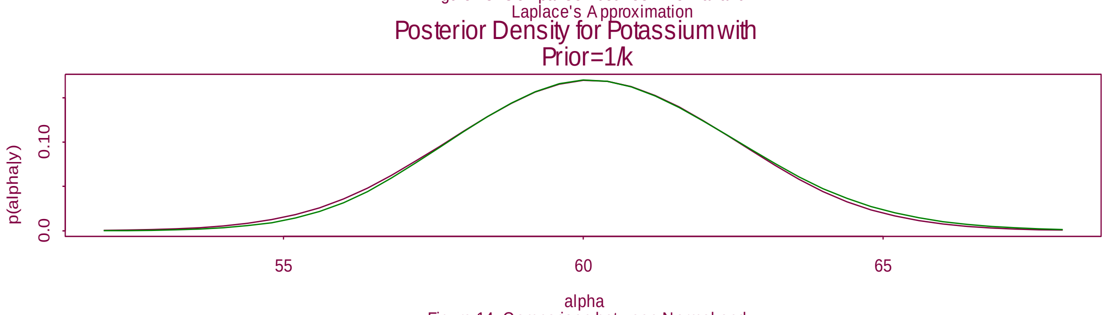


Figure 14: Comparison between Normal and Laplace's Approximation

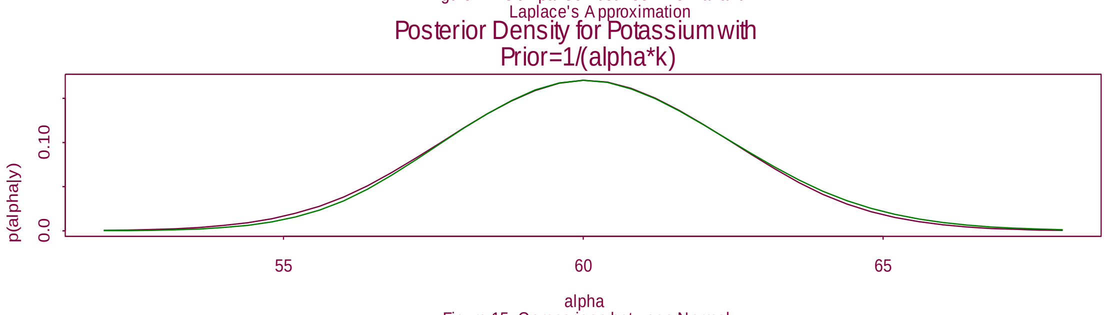


Figure 15: Comparison between Normal and Laplace's Approximation

Comparing Normal and Laplace's Approximation of k of Gamma Distribution with different priors using S-PLUS and R.

```
# Comparing Normal and Laplace's Approximation of k of Gamma Distribution

  with different priors.

# S-PLUS and R.

Norm.Lap<-function(x)

{

n<-length(x)

k<-seq(2.5,3.7,length=13)

plot(spline(b,dnorm(b,mean=3.070154,sd=0.1101784),n=5*length(b),

      xmin=min(b),xmax=max(b)),xlab="b",ylab="p(k|x)",main="Posterior
```

```
    Density    for    Potassium    with    Prior=1",sub="Figure   3.16:
   Comparison

    between Normal and Laplace's Approximation",type="l",col=3)

dalpha<-c(0.646208,0.726896,0.814036,0.907876,1.008664,1.116648,

          1.232073,1.355196,1.486258,1.625509,1.773195,1.929566,

          2.094865)

Lest<-8382.265

Lestalpha<-c(8389.725,8392.500,8388.448,8385.477,8383.477,8382.471,

             8382.301,8382.941,8384.339,8386.446,8389.222,8392.626,

             8396.622)

pb<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha)*exp(-(Lestalpha-Lest))

lines(spline(k,pb,n=5*length(b),xmin=min(b),xmax=max(b)),col=4)

plot(spline(b,dnorm(b,mean=3.06621,sd=0.1101011),n=5*length(b),

    xmin=min(b),xmax=max(b)),xlab="b",ylab="p(b|x)",main="Posterior

    Density for Potassium with Prior=1/b",sub="Figure 3.17:

    Comparison       between       Normal       and       Laplace's
   Approximation",type="l",

    col=3)

dalpha1<-c(0.646208,0.726896,0.814036,0.907876,1.008664,1.116648,

           1.232073,1.355196,1.486258,1.625509,1.773195,1.929566,

           2.094865)

 Lest1<-8383.386

 Lestalpha1<-
c(8398.641,8393.456,8389.441,8386.507,8384.573,8383.570,


8383.433,8384.104,8385.532,8387.670,8390.475,8393.907,

               8397.930)
```

```
 pk1<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha1)*exp(-(Lestalpha1-Lest1))

 lines(spline(b,pb1,n=5*length(b),xmin=min(b),xmax=max(b)),col=4)

plot(spline(b,dnorm(b,mean=3.070153,sd=0.1101783),n=5*length(b),

     xmin=min(b),xmax=max(b)),xlab="b",ylab="p(b|x)",main="Posterior

   Density  for  Potassium  with  Prior=1/(alpha*b)",sub="Figure  3.18:
   Comparison between Normal and Laplace's Approximation", type="l",

   col=3)

dalpha2<-c(0.646760,0.727492,0.844679,0.908567,1.009405,1.117441,

            1.232920,1.356099,1.487218,1.626528,1.774276,1.930709,

            2.096072)

 Lest2<-8387.482

 Lestalpha2<-
c(8402.942,8397.717,8393.665,8390.694,8388.726,8387.689,

            8387.519,8388.158,8389.556,8391.663,8394.439,8397.843,

            8401.840)

 pb2<-1/sqrt(2*pi)*sqrt( 98.59129/dalpha2)*exp(-(Lestalpha2-Lest2))

 lines(spline(b,pb2,n=5*length(b),xmin=min(b),xmax=max(b)),col=4)

}

Norm.Lap(dbmdata$Potassium)

leg.names<-c("Normal Approximation","Laplace's Approximation")

legend(locator(1),leg.names,fill=3:4)
```

Posterior Density for Potassium with Prior=1

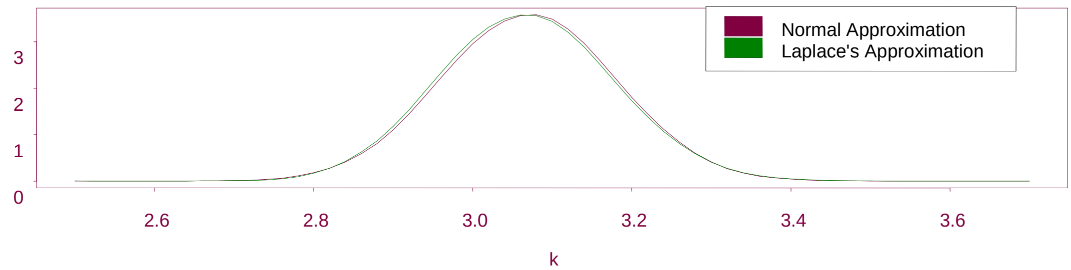


Figure 3.16: Comparison between Normal and Laplace's Approximation

Posterior Density for Potassium with Prior=1/k

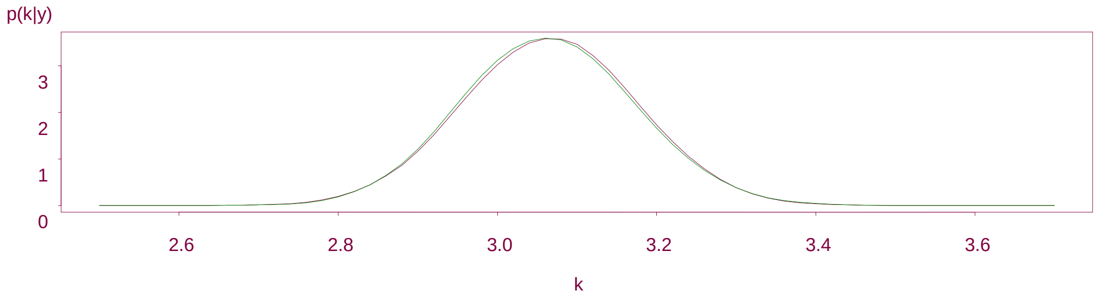


Figure 3.17:Comparison between Normal and Laplace's Approximation

Posterior Density for Potassium with Prior=1/(alpha*k)

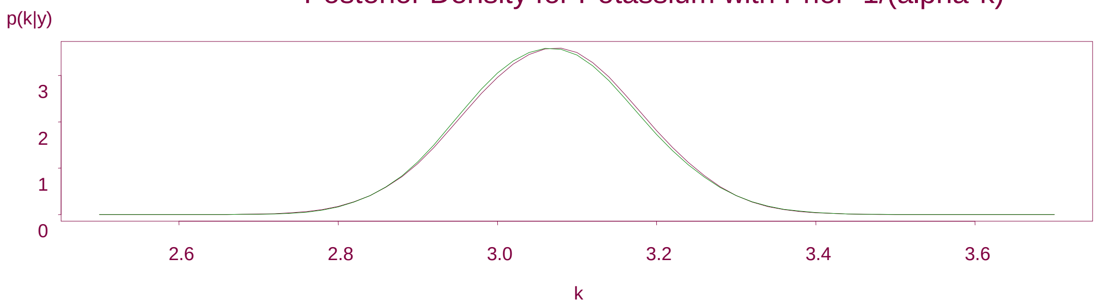


Figure 3.18: Comparison between Normal and Laplace's Approximation

p(k|y)

CHAPTER – 4

POSTERIOR APPROXIMATIONS TO NORMAL DISTRIBUTION

4.1 Introduction:

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 $\Phi(x)$ .The work of Gauss in 1809, 1816 respectively established techniques based on the normal distribution which became standard methods used during the nineteenth century. Davir (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

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

with mean μ and variance σ^2.

4.2 Maximum likelihood estimate of normal distribution:-

Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample of size n from normal population with pdf

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

And the likelihood function is given as

$$L(x \mid \mu, \theta) = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\} \right]$$

$$L(x \mid \mu,\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$$

The log likelihood is given as

$$LogL(x \mid \mu,\theta) = \frac{-n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Case1: when σ^2 is known, the likelihood equation for estimating μ is:

$$\frac{\partial \log L}{\partial\sigma^2} = 0$$

$$\Rightarrow \frac{-1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Case2: when μ is known, the likelihood for estimating σ^2 is

$$\frac{\partial}{\partial\sigma^2}\log L(x \mid \mu,\theta) = 0$$

$$\Rightarrow \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Case 3: Both unknown: The likelihood equation for simultaneous estimation of μ and σ^2 are;

$$\frac{\partial}{\partial\mu}\log L(x \mid \mu,\theta) = 0 \quad and \quad \frac{\partial}{\partial\sigma^2}\log L(x \mid \mu,\theta) = 0$$

Thus giving $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2, sample\,variance$

4.3 Bayesian Estimation for the Parameters of Normal distribution:

Consider two parameter normal distribution

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

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 $g(\mu,\sigma) \propto \frac{1}{\sigma}$, $\sigma > 0$ or class of priors $g(\mu, \sigma \mid c) \propto \frac{1}{\sigma^c}$, $c > 0$ which we consider here.

The likelihood function is given by

$$L(x \mid \mu, \theta) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma)$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right\}$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right]\right\}$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2} \left[A + n(\bar{x} - \mu)^2\right]\right\}, \qquad where\ A = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Posterior is given by

$$P(\mu, \sigma^2 \mid x) \propto g(\sigma)\ L(x \mid \mu, \theta)$$

$$\propto \frac{1}{\sigma^c} \cdot \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2} \left[A + n(\bar{x} - \mu)^2\right]\right\}$$

$$= k \frac{1}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2} \left[A + n(\bar{x} - \mu)^2\right]\right\} \qquad (4.3.1)$$

where $K^{-1} = \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2} \left[A + n(\bar{x} - \mu)^2\right]\right\} d\mu\, d\sigma$

$$= \int_0^\infty \frac{\exp\left(-A/2\sigma^2\right)}{\sigma^{n+c}} \left\{ \int_{--\infty}^{\infty} \exp\left[\frac{-n(x-\mu)^2}{2\sigma^2}\right] d\mu \right\} d\sigma$$

$$= \int_0^\infty \frac{\exp\left(-A/2\sigma^2\right)}{\sigma^{n+c}} \left\{ \frac{\sqrt{2\pi}\sigma}{\sqrt{n}} \right\} d\sigma$$

$$= \frac{\sqrt{2\pi}}{\sqrt{n}} \int_0^\infty \frac{\exp\left(-A/2\sigma^2\right)}{\sigma^{n+c-1}} d\sigma$$

$$= \frac{\sqrt{2\pi}}{\sqrt{n}} \int_0^\infty \frac{\exp\left(-A/2\sigma^2\right)}{(2\sigma)^{\frac{1}{2}(n+c-1)}} (2)^{\frac{1}{2}(n+c-1)} d\sigma$$

$$Put \quad 2\sigma^2 = k \;\Rightarrow\; \sigma = \frac{\sqrt{k}}{\sqrt{2}} \Rightarrow 4\sigma\, d\sigma = dk \;\; or \;\; d\sigma = \frac{\sqrt{2}\, dk}{4\sqrt{k}}$$

$$\Rightarrow \quad K^{-1} = \frac{\sqrt{2\pi}}{\sqrt{n}} \int_0^\infty \frac{\exp(-A/k)\; 2^{\frac{1}{2}(n+c-1)}}{k^{\frac{1}{2}(n+c-1)}} \frac{\sqrt{2}}{4\sqrt{k}}\, dk$$

$$= \sqrt{\frac{2\pi}{n}}\; 2^{\frac{1}{2}(n+c)-2} \int_0^\infty \frac{\exp(-A/k)}{k^{\frac{1}{2}(n+c-1)+1}}\, dk$$

$$= \sqrt{\frac{2\pi}{n}}\; 2^{\frac{1}{2}(n+c-4)} \frac{\Gamma\left(n+c-2\right)/2}{A^{\frac{n+c-2}{2}}}$$

$$\Rightarrow \quad K = \sqrt{\frac{n}{2\pi}} \; \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}}$$

$$\therefore \quad P\left(\mu, \sigma^2 \mid x\right) = \sqrt{\frac{n}{2\pi}} \; \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \; \frac{1}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{x}-\mu)^2\right]\right\}$$

(4.3.2)

The marginal posterior of σ is given by integrating out μ in (4.3.1) we have

$$P(\sigma \mid x) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c}} \int_{-\infty}^{\infty} \exp\left\{\frac{-n}{2\sigma^2}(x-\mu)^2\right\} d\mu$$

$$=\sqrt{\frac{n}{2\pi}}\;\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\;\frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c}}\;\frac{\sqrt{2\pi}\;\sigma}{\sqrt{n}}$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\;\frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}}$$

(4.3.3)

Bayes estimator of σ is given by

$$\hat{\sigma}=E\left(\sigma\mid x\right)=\int_0^\infty \sigma\;P(\sigma\mid x)\;d\sigma$$

$$=\int_0^\infty \sigma\;\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\;\frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}}\;d\sigma$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-2}}\;d\sigma$$

$$=\frac{A^{\frac{n+c-2}{2}}\;2^{\frac{1}{2}(n+c-2)}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{(2\sigma^2)^{(n+c-2)/2}}\;d\sigma$$

substituting $2\sigma^2=z$, we have

$$\hat{\sigma}=\frac{A^{\frac{n+c-2}{2}}\;2^{\frac{1}{2}(n+c-2)-\frac{3}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}(n+c-4)}}\int_0^\infty \frac{\exp\left\{-A/z\right\}}{(z)^{\frac{(n+c-3)}{2}+1}}\;dz$$

$$=\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)\,2^{\frac{1}{2}}}\;\frac{\Gamma\left(n+c-3/2\right)}{A^{\frac{n+c-3}{2}}}$$

$$=\sqrt{\frac{A}{2}}\;\frac{\Gamma\left(\frac{n+c-3}{2}\right)}{\Gamma\left(\frac{n+c-2}{2}\right)} \qquad (4.3.4)$$

Bayes estimator of σ^2 is

$$\hat{\sigma}^2 = E\left(\sigma^2 \mid x\right) = \int_0^\infty \sigma^2 \; P(\sigma \mid x)\, d\sigma$$

$$= \int_0^\infty \sigma^2 \; \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-1}} \; d\sigma$$

$$= \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{\sigma^{n+c-3}} \; d\sigma$$

$$= \frac{A^{\frac{n+c-2}{2}} \; 2^{\frac{1}{2}(n+c-3)}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/2\sigma^2\right\}}{(2\sigma^2)^{(n+c-3)/2}} \; d\sigma$$

substituting $2\sigma^2 = t$, *we have*

$$\hat{\sigma}^2 = \frac{A^{\frac{n+c-2}{2}} \; 2^{\frac{1}{2}(n+c-3)}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left\{-A/t\right\}}{(t)^{\frac{(n+c-4)}{2}+1}} \frac{\sqrt{2}}{4\sqrt{t}} dt$$

$$= \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2} \frac{\Gamma\left(n+c-4/2\right)}{A^{\frac{n+c-4}{2}}}$$

$$= \frac{A}{n+c-4} \qquad (4.3.5)$$

If we put c=4 in (4.3.5) 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.

Marginal of μ is given by

$$P(\mu \mid x) = \sqrt{\frac{n}{2\pi}} \frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right) 2^{\frac{1}{2}(n+c-4)}} \int_0^\infty \frac{\exp\left[\frac{-1}{2\sigma^2}\left\{A + n(\bar{x}-\mu)^2\right\}\right]}{\sigma^{n+c}} d\sigma$$

$$=\int_{-\infty}^{\infty}\mu\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}d\mu$$

$$=\sqrt{\frac{n}{2\pi}}\frac{A^{\frac{n+c-2}{2}}\,2^{\frac{n+c}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)2^{\frac{1}{2}(n+c-4)}}\int_0^{\infty}\frac{\exp\left[\frac{-1}{2\sigma^2}\left\{A+n(\bar{x}-\mu)^2\right\}\right]}{(2\sigma^2)^{(n+c)/2}}d\sigma$$

Put $2\sigma^2=r$, we have

$$P(\mu\mid x)=\sqrt{\frac{n}{2\pi}}\frac{A^{\frac{n+c-2}{2}}\sqrt{2}}{\Gamma\left(\frac{n+c-2}{2}\right)}\int_0^{\infty}\frac{\exp\left[\frac{-1}{r}\left\{A+n(\bar{x}-\mu)^2\right\}\right]}{(r)^{\frac{n+c-1}{2}+1}}dr$$

$$=\sqrt{\frac{n}{\pi}}\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)}\frac{\Gamma\left(\frac{n+c-1}{2}\right)}{\left\{A+n(\bar{x}-\mu)^2\right\}^{\frac{n+c-1}{2}}}$$

$$=\sqrt{\frac{n}{\pi}}\frac{A^{\frac{n+c-2}{2}}}{\Gamma\left(\frac{n+c-2}{2}\right)}\frac{\Gamma\left(\frac{n+c-1}{2}\right)}{A^{\frac{n+c-1}{2}}\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}$$

$$=\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}:\quad -\infty<\mu<\infty \tag{4.3.6}$$

Bayes estimator of μ is given by

$$\hat{\mu}=E(\mu\mid x)=\int_{-\infty}^{\infty}\mu\,P(\mu\mid x)d\mu$$

$$=\sqrt{\frac{n}{A}}\frac{1}{\beta\left(\frac{1}{2},\frac{n+c-2}{2}\right)}\int_{-\infty}^{\infty}\frac{\mu}{\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}}d\mu$$

$$\textit{substituting} \quad \frac{n(\bar{x}-\mu)^2}{A} = \frac{t^2}{n+c-2} \quad \textit{we have}$$

$$\hat{\mu} = \sqrt{\frac{n}{A}} \frac{1}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)} \int_{-\infty}^{\infty} \frac{\bar{x} - \frac{\sqrt{A/n}\; t}{\sqrt{n+c-2}}}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} \left(\frac{-\sqrt{A/n}}{\sqrt{n+c-2}}\right) dt$$

$$= \frac{-1}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{\bar{x} - \frac{\sqrt{A/n}\; t}{\sqrt{n+c-2}}}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} dt$$

$$= \frac{-\bar{x}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{1}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} dt + \frac{\sqrt{A/n}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)} \int_{-\infty}^{\infty} \frac{t}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} dt$$

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.

$$\therefore \hat{\mu} = \frac{-\bar{x}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{1}{\left\{1+\frac{t^2}{n+c-2}\right\}^{\frac{n+c-1}{2}}} dt$$

$$= \frac{-\bar{x}}{\beta\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{-\sqrt{n+c-2}\,\sqrt{A/n}}{\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{\frac{n+c-1}{2}}} d\mu = \bar{x} \tag{4.3.7}$$

4.4 Bayesian intervals for parameter of normal distribution:

The joint posterior of μ and σ^2 is given by

$$P(\mu, \sigma \mid x) = \frac{k}{\sigma^{n+c}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{x}-\mu)^2\right]\right\}; \; -\infty < \mu < \infty; \sigma > 0 \tag{4.4.1}$$

where k is normalizing constant.

Putting c=2 in the (4.4.1) we have

$$P(\mu, \sigma^2 \mid x) = \frac{k}{\sigma^{n+2}} \exp\left\{\frac{-1}{2\sigma^2}\left[A + n(\bar{x}-\mu)^2\right]\right\}$$

Where $$A=\sum_{i=1}^{n}(x_i-\bar{x})^2 \; and \; k^{-1}=\sqrt{\frac{\pi}{n}}\left(\frac{2}{A}\right)^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right) \qquad (4.4.2)$$

Integrating out μ and restoring the normalizing constant k, the marginal posterior density for σ^2 is given by

$$P\left(\sigma^2 \mid x\right)=\left(\frac{A}{2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\exp\left(-A/2\sigma^2\right)}{\left(\sigma^2\right)^{(n-1/2)+1}}$$

(4.4.3)

Similarly we obtain the marginal posterior of μ

$$P(\mu \mid x)=\frac{\sqrt{n}}{\sqrt{A}\,\beta\left(\frac{1}{2},\frac{n-1}{2}\right)} \frac{1}{\left[1+\frac{n(\bar{x}\_\mu)^2}{A}\right]^{\frac{n}{2}}}$$

(4.4.4)

from (4.8.3) it follows A/σ^2 is distributed as χ^2 with (n-1) degrees of freedom

4.5 Normal Approximation for normal distribution:

The pdf of normal distribution is given by

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

The likelihood function is given by

$$L\left(\mu,\sigma^2 \mid x\right)=\frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}}\exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\right]$$

Consider the prior $g\left(\mu,\sigma^2\right)=1$

Therefore posterior density is given by

$$P\left(\mu,\sigma^2 \mid x\right) \propto p\left(\mu,\sigma^2\right) L\left(\mu,\sigma^2 \mid x\right)$$

$$\propto \frac{1}{(\sigma^2)^{n/2}(2\pi)^{1/2}} \exp\left[\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\right]$$

$$\log P(\mu,\sigma^2 \mid x) = \log constant - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

(4.5.1)

differentating $= n$ (4.5.1) *partially w.r.t* μ, σ^2 *we have*

$$\frac{\partial \log P(\mu,\sigma^2 \mid x)}{\partial\mu} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}$$

$$\frac{\partial \log P(\mu,\sigma^2 \mid x)}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}$$

Posterior mode is $\hat{\mu} = \bar{x}$ *and* $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{(n-1)s^2}{n}$

$$\frac{\partial^2 \log(\mu,\sigma^2 \mid x)}{\partial\mu^T\,\partial\mu} = \frac{-n}{\sigma^2}$$

$$or\ \frac{-\partial^2 \log(\mu,\sigma^2 \mid x)}{\partial\mu^T\,\partial\mu} = \frac{n}{\sigma^2}$$

$$\frac{\partial^2 \log(\mu,\sigma^2 \mid x)}{\partial{\sigma^2}^T\,\partial\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(\bar{x}-\mu)^2 = \frac{n}{2\sigma^4} - \frac{n\sigma^2}{2\sigma^4} = \frac{-n}{2\sigma^4}$$

$$\therefore \frac{-\partial^2 \log(\mu,\sigma^2 \mid x)}{\partial{\sigma^2}^T\,\partial\sigma^2} = \frac{n}{2\sigma^4}$$

$$\therefore I(\mu,\sigma^2) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$

$$\therefore I^{-1}(\mu,\sigma^2) = \frac{1}{n^2/2\sigma^6}\begin{pmatrix} n/2\sigma^4 & 0 \\ 0 & n/\sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}$$

$$\therefore I^{-1}(\hat{\mu},\hat{\sigma}^2) = \begin{pmatrix} \hat{\sigma}^2/n & 0 \\ 0 & 2\hat{\sigma}^4/n \end{pmatrix}$$

4.6 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 $x_1, x_2, ..., x_n$ 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 $X : x_1, x_2, ..., x_n$ is

$$L(X:\sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-1}{2\sigma^2}(x_i - \mu)^2 \right\} \right]$$

$$L(X:\sigma^2) = \left(\frac{1}{\sigma^2 2\pi} \right)^{n/2} \exp\left\{ \frac{-w}{2\sigma^2} \right\}$$

(4.6.1)

$$where\ w = \sum_{i=1}^{n} (x_i - \mu)^2$$

4.7 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:

$$f_1(\sigma^2) = \frac{b_1^{\frac{a_1}{2}}}{2^{\frac{a_1}{2}}\Gamma\left(\frac{a_1}{2}\right)}(\sigma^2)^{\frac{-a_1}{2}-1}\exp\left\{\frac{-b_1}{2\sigma^2}\right\}, \sigma^2 > 0; a_1, b_1 > 0 \tag{4.7.1}$$

The density kernel is

$$P_1(\sigma^2) \propto (\sigma^2)^{\frac{-a_1}{2}-1}\exp\left\{\frac{-b_1}{2\sigma^2}\right\}$$

(4.7.2)

Now the posterior distribution of the parameter σ^2 for the given data $X : x_1, x_2, ...,x_n$ is

$$P_1(\sigma^2 \mid X) \propto L(X;\sigma^2)P_1(\sigma^2)$$

$$P_1(\sigma^2 \mid X) \propto (\sigma^2)^{\frac{-n}{2}} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{\frac{-a_1}{2}-1} e^{\frac{-b_1}{2\sigma^2}}$$

$$P_1(\sigma^2 \mid X) \propto (\sigma^2)^{\frac{-\alpha_1}{2}-1} e^{\frac{-\beta_1}{2\sigma^2}}$$

(4.7.3)

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.

4.8 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

$$f_2(\sigma^2) = \frac{b_2^{\,a_2}}{\Gamma(a_2)}(\sigma^2)^{-(a_2+1)}\exp\left\{\frac{-b_1}{\sigma^2}\right\}, \sigma^2 > 0; a_2, b_2 > 0$$

(4.8.1)

Now the posterior distribution of the parameter σ^2 for given data $X : x_1, x_2, ..., x_n$ is

$$P_2(\sigma^2 \mid X) \propto (\sigma^2)^{-(a_2+1)} e^{\frac{-b_2}{\sigma^2}} (\sigma^2)^{\frac{-n}{2}} e^{\frac{-w}{2\sigma^2}}$$

$$P_2(\sigma^2 \mid X) \propto (\sigma^2)^{-\left(a_2+\frac{n}{2}+1\right)} e^{\frac{-1}{\sigma^2}\left(\frac{2b_2+w}{2}\right)}$$

$$P_2(\sigma^2 \mid X) \propto (\sigma^2)^{-(\alpha_2+1)} e^{\frac{-\beta_2}{\sigma^2}}$$

(4.8.2)

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.

4.9 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

$$f_3(\sigma^2) = \sqrt{\frac{b_3}{2\pi}}(\sigma^2)^{\frac{-3}{2}} e^{\frac{-b_3}{2\sigma^2}} ; \sigma^2 > 0, b_3 > 0$$

(4.9.1)

Now the posterior distribution of the parameter σ^2 for given data $X : x_1, x_2, ..., x_n$ is

$$P_3(\sigma^2 \mid X) \propto (\sigma^2)^{-\left(\frac{n}{2}\right)} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{-\left(\frac{3}{2}\right)} e^{\frac{-b_3}{2\sigma^2}}$$

$$P_3(\sigma^2 \mid X) \propto (\sigma^2)^{-(\alpha_3+1)} e^{\frac{-\beta_3}{\sigma^2}}$$

(4.9.2)

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.

4.10 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:

$$f_4(\sigma^2) = a_4 b_4 (\sigma^2)^{-(a_4+1)} e^{-b_4(\sigma^2)^{-a_4}} \;\; ; \;\; where\ \sigma^2 > 0,\ a_4, b_4 > 0$$

For making the conjugate prior, we take $a_4 = 1$ then the prior is:

$$f_4(\sigma^2) = b_4 (\sigma^2)^{-2} e^{-b_4(\sigma^2)^{-1}} \tag{4.10.1}$$

Now the posterior distribution of the parameter σ^2 for given data $X : x_1, x_2, \ldots.., x_n$ is

$$P_4(\sigma^2 \mid X) \propto (\sigma^2)^{-\left(\frac{n}{2}\right)} e^{\frac{-w}{2\sigma^2}} (\sigma^2)^{-2} e^{\frac{-b_4}{\sigma^2}}$$
(4.10.2)

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.

4.11 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 $Y = X_{n+1}$ given that $X : x_1, x_2, \ldots.., x_n$

under posterior inverse chi-squared distribution is:

$$P_1(Y \mid X) = \int_0^\infty \phi(y \mid \sigma^2)\, P_1(\sigma^2 \mid X)\, d\sigma^2$$

$$P_1(Y \mid X) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \frac{\beta_1^{\alpha_1/2} (\sigma^2)^{\frac{-\alpha_1}{2}-1}}{2^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1}{2}\right)} \exp\left\{\frac{-\beta_1}{2\sigma^2}\right\} d\sigma^2$$

$$P_1(Y \mid X) = \frac{\beta_1^{\alpha_1/2}}{2^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1}{2}\right)(2\pi)^{1/2}} \int_0^\infty \frac{1}{(\sigma^2)^{\frac{\alpha_1}{2}+\frac{1}{2}+1}} \exp\left\{\frac{-1}{\sigma^2}\left(\frac{(y-\mu)^2}{2} + \frac{\beta_1}{2}\right)\right\} d\sigma^2$$

$$P_1(Y \mid X) = \frac{\beta_1^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1+1}{2}\right)}{2^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1}{2}\right)(2\pi)^{1/2}\left[\frac{\beta_1}{2} + \frac{(y-\mu)^2}{2}\right]^{\frac{\alpha_1+1}{2}}}$$

$$P_1(Y \mid X) = \frac{\beta_1^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1+1}{2}\right) 2^{\frac{\alpha_1+1}{2}} \left[1 + \frac{(y-\mu)^2}{\alpha_1(\beta_1/\alpha_1)}\right]^{\frac{-(\alpha_1+1)}{2}}}{2^{\alpha_1/2}\,\Gamma\left(\frac{\alpha_1}{2}\right)(2\pi)^{1/2}\,\beta_1^{\frac{\alpha_1+1}{2}}}$$

$$P_1(Y \mid X) = \frac{\Gamma\left(\frac{\alpha_1+1}{2}\right)\left[1 + \frac{(y-\mu)^2}{\alpha_1(\beta_1/\alpha_1)}\right]^{\frac{-(\alpha_1+1)}{2}}}{\sqrt{\frac{\beta_1}{\alpha_1}}\sqrt{\alpha_1\pi}\,\Gamma\left(\frac{\alpha_1}{2}\right)}$$

(4.11.1)

which is the probability density function of t-distribution i.e.

$$Y \mid X \sim t\left(\alpha_1, \mu, \frac{\beta_1}{\alpha_1}\right); -\infty < x, \mu < \infty; \alpha_1 > 0$$

(4.11.2)

Hence $\mathrm{Y|X}$ has the t-distribution with three parameters $\mathrm{u_1, v_{1,}}$ and $\mathrm{w_1}$

Where $u_1 = \alpha_1,\ v_2 = \mu$ *and* $w_2 = \dfrac{\beta_1}{\alpha_1}$; $w_1 > 0$

b)The Posterior Predictive Distribution under the Prior Inverted Gamma Distribution:

The posterior predictive distribution for $Y = X_{n+1}$ given that $X : x_1, x_2, \ldots\ldots, x_n$

under posterior inverted gamma distribution is:

$$P_2(Y \mid X) = \frac{\Gamma\left(\alpha_2 + \frac{1}{2}\right)\left[1 + \frac{(x-\mu)^2}{2\alpha_1(\beta_2/\alpha_2)}\right]^{\frac{-(2\alpha_2+1)}{2}}}{\sqrt{(\beta_2/\alpha_2)}\sqrt{2\alpha_2\pi}\,\Gamma\left(\frac{2\alpha_2}{2}\right)}$$

(4.11.3)

which is the probability density function of t-distribution i.e.

$$\mathrm{Y} \mid \mathrm{X} \sim \mathrm{t}\left(2\alpha_1, \mu, \frac{\beta_2}{\alpha_{21}}\right);\ -\infty < \mathrm{x}, \mu < \infty\ ; \alpha_2 > 0$$

(4.11.4)

Hence $\mathrm{Y} \mid \mathrm{X}$ has the t-distribution with three parameters $u_2, v_2,\ and\ w_2$

where $u_2 = 2\alpha_2,\ v_1 = \mu\ and\ w_1 = \frac{\beta_2}{\alpha_2}\ ;\ w_2 > 0$.

c)The Posterior Predictive Distribution under the Prior Levy Distribution:

The posterior predictive distribution for $Y = X_{n+1}$ given that $X : x_1, x_2, ...,x_n$

under posterior inverted gamma distribution is:

$$P_3(Y \mid X) = \int_0^\infty P(y \mid \sigma^2)\, P_3(\sigma^2 \mid X)\, d\sigma^2$$

$$P_3(Y \mid X) = \frac{\Gamma\left(\alpha_3 + \frac{1}{2}\right)\left[1 + \frac{(y-\mu)^2}{2\alpha_3(\beta_3/\alpha_3)}\right]^{\frac{-(2\alpha_3+1)}{2}}}{\sqrt{(\beta_3/\alpha_3)}\sqrt{2\alpha_3\pi}\,\Gamma\left(\frac{2\alpha_2}{2}\right)}$$

(4.11.5)

Which is the probability density function of t-distribution i.e.

$$\mathrm{Y} \mid \mathrm{X} \sim \mathrm{t}\left(2\alpha_3, \mu, \frac{\beta_3}{\alpha_3}\right);\ -\infty < \mathrm{x}, \mu < \infty\ ; \alpha_3 > 0 \qquad (4.11.6)$$

Hence $\mathrm{Y} \mid \mathrm{X}$ has the t-distribution with three parameters $\mathrm{u}_3, \mathrm{v}_3,$ and w_3

where $u_3 = 2\alpha_3,\ v_3 = \mu$ and $w_3 = \dfrac{\beta_3}{\alpha_3}$; $w_3 > 0$.

d)The Posterior Predictive Distribution under the Prior Gumbel Type-II Distribution:

The posterior predictive distribution for $Y = X_{n+1}$ given that $X : x_1, x_2, \ldots, x_n$

under posterior inverted gamma distribution is:

$$P_3(Y \mid X) = \frac{\Gamma\left(\alpha_4 + \frac{1}{2}\right)\left[1 + \frac{(y-\mu)^2}{2\alpha_4(\beta_4/\alpha_4)}\right]^{\frac{-(2\alpha_4+1)}{2}}}{\sqrt{(\beta_4/\alpha_4)}\sqrt{2\alpha_4\pi}\,\Gamma\left(\frac{2\alpha_4}{2}\right)}$$

(4.11.7)

$$Y \mid X \sim t\left(2\alpha_4, \mu, \frac{\beta_4}{\alpha_4}\right); \qquad -\infty < x, \mu < \infty\ ; \alpha_4 > 0$$

Hence $Y \mid X$ has the t-distribution with three parameters $u_4, v_4,$ *and* w_4

Where $u_4 = 2\alpha_4,\ v_4 = \mu$ *and* $w_4 = \dfrac{\beta_4}{\alpha_4}$; $w_4 > 0$

4.12 Comparison of priors with respect to posterior variances:

The variances of the posterior distributions are calculated and are given in Table 4.2.1, 4.2.2, 4.2.3:

1. For the posterior inverse chi-square distribution we have

$$V(\theta \mid Y) = \frac{2\beta_1^{\ 2}}{(\alpha_1 - 2)^2(\alpha_1 - 4)}\text{;provided}\,\alpha_1 > 4$$

2. For the posterior inverted gamma distribution we have

$$V(\theta \mid Y) = \frac{\beta_i^{\ 2}}{(\alpha_i - 1)^2(\alpha_i - 2)} \qquad \text{provided } \alpha_i > 2, i = ,2,3,4$$

4.13 Comparison using the posterior predictive variances:

The posterior predictive variances using different prior distributions are given in the tables 4.2.4, 4.2.5 and 4.2.6.

The posterior predictive variances under inverse chi-square as prior distribution is

$$V(\theta \mid Y) = \frac{\beta_1}{\alpha_1 - 2}; \alpha_1 > 2$$

and the posterior predictive variances under the inverted gamma, Levy and Gumbel typeII distributions as priors is

$$V(\theta \mid Y) = \frac{2\beta_i}{\alpha_i - 2}; \alpha_i > 2 : i = 2,3,4$$

Example 4.1: Simon new comb 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 4.1.1.

#Bayesian analysis of normal distribution with different priors in R.
#Prior=1.

```
pos.normal<-function(theta,x)
{
z<-(x-theta[1])/theta[2]
n<-length(x)

lik<- n*log(theta[2])+sum(z^2)

pri<--log(1)

pos<-pri+lik

return(pos)

}

speed<-
c(28,29,24,37,36,26,29,26,22,20,25,23,32,27,33,24,36,28,27,32,
28,24,21,32,26,27,24,29,34,25,36,30,28,39,16,-
44,30,28,32,27,28,23,27,23,25,36,31,
24,16,29,21,26,27,25,40,31,28,30,26,32,-2,19,29,22,33,25)

out<-nlm(pos.normal,x=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,x)
{
z<-(x-theta[1])/theta[2]
n<-length(x)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/theta[2])
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,x=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,x)
{
z<-(x-theta[1])/theta[2]
n<-length(x)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^2))
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,x=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,x)
{
```

```
z<-(x-theta[1])/theta[2]
n<-length(x)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^3))
pos<-pri+lik
return(pos)
}
out<-nlm(pos.normal,x=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,x)
{
z<-(x-theta[1])/theta[2]
n<-length(x)
lik<- n*log(theta[2])+sum(z^2)
pri<--log(1/(theta[2]^4))
pos<-pri+lik
return(pos)
```

```
}
out<-nlm(pos.normal,x=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 4.1.1: 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: 4.2 (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,x){
```

```
n<-length(x); w<-sum((x-mu)^2)

alpha1<-(a1+n);beta1<-b1+w

pvc<-2*(beta1^2)/(((alpha1-2)^2)*(alpha1-4))

return(pvc)

 }

a1=b1=5
mu<-20
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))

x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(sim.var(a1,b1,mu,x1),sim.var(a1,b1,mu,x2),sim.var(a
1,b1,mu,x3),sim.var (a1,b1,mu,x4))
```

Posterior variance of sigma^2 under inverted gamma as a prior

```
sim.var <-function(a2,b2,mu,x){
n<-length(x); w<-sum((x-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
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))

x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(sim.var(a2,b2,mu,x1),sim.var(a2,b2,mu,x2),sim.var(a
2,b2,mu,x3),sim.var (a2,b2,mu,x4))
```

Posterior variance of sigma^2 under levy distribution as a prior

```
sim.var <-function(a3,b3,mu,x){
n<-length(x); w<-sum((x-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
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))
```

```
x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(sim.var(a3,b3,mu,x1),sim.var(a3,b3,mu,x2),sim.var(a3,b3,mu,x3),sim.var (a2,b2,mu,x4))
```

Posterior variance of sigma^2 under Gumbel type II distribution as a prior

```
sim.var <-function(a4,b4,mu,x){

n<-length(x); w<-sum((x-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
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))

x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(sim.var(a4,b4,mu,x1),sim.var(a4,b4,mu,x2),sim.var(a4,b4,mu,x3),sim.var (a4,b4,mu,x4))
```

The results obtained using above programme are presented in tables 4.2.1 : 4.2.2; 4.2.3 for different values of hyper parameters, n and mean.

Table 4.2.1: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
		5	0.94822	0.66542	1.27028	1.37267
		10	0.66542	0.37790	1.37267	1.58936

60	V2	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 4.2.2: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 4.2.3: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
		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

60	**V2**	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 4.2.1, 4.2.2; 4.2.3 for different values of hyper parameters, n and mean. In the above Tables 4.2.1, 4.2.2 and 4.2.3, 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,x){

n<-length(x)

w<-sum((x-mu)^2)

alpha1<-(a1+n)

beta1<-b1+w

pvc<-beta1/(alpha1-2)

return(pvc)

 }

a1=b1=5
```

```
mu<-20
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4));x3<-
rnorm(30,20,sqrt(6))

cbind(pre.var(a1,b1,mu,x1),pre.var(a1,b1,mu,x2),pre.var(a
1,b1,mu,x3))
```

Predictive Posterior variance of sigma^2 under inverted gamma as a prior

```
pre.var <-function(a2,b2,mu,x){

n<-length(x)

w<-sum((x-mu)^2)

alpha2<-a2+n/2

beta2<-(2*b2+w)/2

pvg<-beta2/(alpha2-1)

return(pvg)

 }

a2=b2=5
mu<-20
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4));x3<-
rnorm(30,20,sqrt(6))

cbind(pre.var(a1,b1,mu,x1),pre.var(a1,b1,mu,x2),pre.var(a
1,b1,mu,x3))
```

Predictive Posterior variance of sigma^2 under levy distribution as a prior

```
pre.var <-function(a3,b3,mu,x){

n<-length(x); w<-sum((x-mu)^2)

alpha3<-(1+n/2);beta3<-(b3+w)/2

pvl<-(beta3)/(alpha3-1)

return(pvl)

 }

a3=b3=5
mu<-20
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))
```

```
x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(pre.var(a3,b3,mu,x1),pre.var(a3,b3,mu,x2),pre.var(a3,b3,mu,x3),pre.var (a2,b2,mu,x4))
```

Predictive Posterior variance of sigma^2 under Gumbel type II distribution as a prior

```
pre.var <-function(a4,b4,mu,x){

n<-length(x); w<-sum((x-mu)^2)

alpha4<-(1+n/2);beta4<-(2*b4+w)/2

pvgb<-(beta4)/(alpha4-1)

return(pvgb)

 }

a4=b4=5
mu<-20
x1<-rnorm(30,20,sqrt(2));x2<-rnorm(30,20,sqrt(4))

x3<-rnorm(30,20,sqrt(6));x4<-rnorm(30,20,sqrt(8))

cbind(pre.var(a4,b4,mu,x1),pre.var(a4,b4,mu,x2),pre.var(a4,b4,mu,x3),pre.var (a4,b4,mu,x4))
```

Table: 4.2.4: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
		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

60	**V2**	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 4.2.5: 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
		5	3.053136	2.941562	3.436798	3.638667

60	**V2**	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 4.2.6: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 4.2.4 ; 4.2.5; 4.2.6 for different values of hyper parameters, n and mean. In the above Tables 4.2.4, 4.2.5 and 4.2.6, 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

Abramowitz, M. And Stegun, I.(Ed). (1964), Handbook Of Mathematical Functions, National Bureau Of Standards, Applied Mathematical Series 55,U.S.Government Printing Office.

Ahmed, A.A., Khan, A.A., and Ahmad, S.P.(2007). Bayesian Analysis of Exponential Distribution in S-PLUS and R Softwares. Sri Lankan Journal of Applied Statistics, Vol. 8, 2007, p. 95-109.

Ahmad,S.P., & Bhat,B.A.(2010): Posterior Estimation of Two Parameter Exponential Distribution Using SPLUS Softwares. Journal f Reliability & Statistical Science Vol.3(3), 27-34.

Ahmad,S.P, Ahmad,A., and Khan, A.A., (2011):Bayesian Analysis of Gamma Distribution Using SPLUS & R-softwares. Asian Journal of Mathematics & Statistics, Vol.4(4),Pp.224-233.

Ahmad,S.P.,(2006): Optimiation Tools in Bayesian Data Analysis. P.hd Thesis Submitted to the Department of Statistics University of Kashmir sgr.

Barnard, G.A.(1949): Statistical Inference (With Discussion).J.Roy.Statist.Soc (Ser.B)11,115-139.

Balakrishna,N. And Ma,Y.(1997): Convergence Rates Of Empirical Bayes Estimation And Selection For Exponential Populations With Location Parameters. In Advance In Statistical Decision Theory And Applications. Bbirkhauser, Boston,Pp.65-77.

Bartholomew, D.J.(1957). A problem In Life Testing. Journal Of American Statistical Association. 52, 350-355.

Bayes, Rev. T.R.(763): “A essay towards solving problems in the doctrine of chances”, Pohil.Trans. Roy. Soc., 53,370-418.Reprinted in Biometrika, 48,296-315.

Bazovsky, I. (1961): Reliability Theory And Practice, Prentice Hall, New Jersey.

Berger, J. O. (1985): Statistical Decision Theory And Bayesian Analysis. 2nd Edn. New York: Springer-Verlag.

Berger, J.O. (1982c): Estimation In Continuous Exponential Families: Bayesian Estimation Subject To Risk Restriction And Inadmissibility Results. In Statistical Decision Theory and Related Top ics Iii, S.S. Gupta And J. Berger (Eds.) Academic Press, New York.

Berger,J.O.(1988): Statistical Decision Theory and Bayesian analysis (2nd edition),New York: Springer-Verlag.

Berger, J. O., and Wolpert, R. (1984). The likelihood principle. Harvard, California: institute of mathematical statistics.

Bernardo, J. M. (2005). Intrinsic Point Estimation Of The Normal Variance. New Dehli: Anamay Pub.

Bernardo,J.M And Smith, A. F. M(1994): Bayesian Theory , J.Wiley And Sons, New York

Birnbaum,A.(1962): On The Foundation Of statistical Inference (With Discussion).J.Amer. Statist. Assoc. 57,267-326.

Bolstad, W. M. (2004). Introduction To Bayesian Statistics. New York: John Wiley & Sons, New York.

Box, G.E.P., And Tiao, G.C. (1973). Bayesian Inference In Statistical Analysis. Addison Wesley, Reading.

Carlin, B. and Levis,T. (1996): Bayes and Empirical Bayes Methods For Data Analysis, Chapman and Hall: London.

Carlin, B.P., and Louis, T.A.(2001). Bayes and Empirical Bayes methods for data analysis, Chapman and hall, London.

Chao, M., And Glaser, R.E. (1978). The Exact Distribution Of Bartlett's Test Statistic For Homogeneity Of Variances With Unequal Sample Sizes. Journal Of The American Statistical Association, 73, 422-426.

Clog, C.C., Rubin, D.B., Schenker, N., Schutz, B., and Wideman,L.(1991). Multiple imputation of industry and occupation codes in census public- use samples using Bayesian logistic regression. Journal of the American statistical association 86, 68-78.

Cox, D.R,R and Reid, N.(1987). Parameter Orthogonality and approximate conditional inference. Journal of the royal statistical society, B,49,1-39.

Davis, D. J. (1952): "The Analysis Of Some Failure Data", J.Amer. Statist. Assoc. 47, 113 150.

De Finitte (1970): Teoria delle Probabilitia, Turin: Einandi, English Translation as Theory of probability in 1975, Chichester: Wiley

DiCiccio, T.J.(1986): approximate conditional inference for location families, candian journal statistics 14,5-18

Edwards, W., Lindman, H., and savage, L. J.(1963). Bayesian statistical inference for psychological research. Psychological review 70, 193-242.

Engelhardt, M., And Bain, L.J. (1978). Prediction Intervals For The Weibull Process Technometrics, 20, 167-170.

Epstein, B. (1960). Test for validity of the assumption that the underline distribution of life is exponential. Technometrics 2, 83-101, 167-183.

Epstein, B. And Sobel, M. (1953). Life Testing. Journal Of The American Statistical Association 48, 486-502.

Epstein, B. And Sobel, M. (1954). Some Theorems Relevant To Life Testing From An Exponential Distribution. Annals Of The Mathematical Statistics, 25, 373-381.

Epstein, B. And Sobel, M. (1955). Sequential Life Testing In Exponential Case. Annals Of The Mathematical Statistics, 26, 82-93.

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.

Feigl, P., and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. Biometrics , 21, 826-836.

Freedman, L. S., Spiegel halter, D. J., and Parmar, M. K. V. (1994). The what, why, and how of Bayesian clinical trials monitoring, statistics in medicine, 13, 1371-1383.

Gauss, C. F. (1809): Theoria Motus Corporum Coelestium, Hamsburd: Perths & Besser.(English Translation By C.H. Davis, Published 1857, Boston : Little Brown, Co.).

Gauss, C. F. (1816):Bestimmuing Der Genauigkeit Der Beobachtugen, Zeit-Schrift Astronomi, I, 185-1

Gelfand, A. E. And Smith, A.F.M (1990):Sampling Based Approached To Calculating Marginal Densities. J. Amer. Statist. Assoc. Based 85,398-409. [An Excellent Primer On Simulation-Based Techniques To Numerical Integration In The Contact Of Bayesian Statistics]

Gelman, A., Carlin, J.B., Stern, H.S., And Rubin, D.B. (1995). Bayesian Data Analysis. Chapman And Hall, London.

Geodman, L. A. (1960). A Note On The Estimation Of Variance", Sankhyz, **22**, 221-228.

Gianola,D., ad Fernando,R.L.,(1986): Bayesian Methods in Animal Breeding Theory.J.Annual.Sci.63,217-244.

Govindarajulu, Z.(1964): A Supplement To Mendenhall’s By Bibliography On Life Testing And Related Topics. Journal Of The American Statistical Association, 59,1231-1291.

Gupta, S.S., And Groll, P.A. (1961). Gamma Distribution In Acceptance Sampling Based On Life Tests. Journal Of The American Statistical Association, 56, 942-970.

Hartigan, J. A.(1965): The asymptotic unbiased prior. Annals of mathematical statistics, 36, 1137- 1152.

Jamali,A.S.,L.J.Lin and D. Yingzhuo,(2006): Effect of Scale Parameters in the Performance Of Shewhrat Control Chart with Interpretation Rules.J.Applied Sci., 6: 2676-2678.

Jeffrey, H.(1939): theory of probability . Oxford University Press.

Jeffrey, H.(1961): Theory Of Probability.III edition(I edition 1939, II edition 1994). Clarendon press ,Oxford University Press.

Johnson, N. L. And Kotz, S. (1970). Continuous Univariate Distributions, Vol. 1 And 2. Houghton-Mifflin, Boston.

Johnson, N.L., Kotz, S., And Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1. John Wiley & Sons, New York.

Johnson, N.L., Kotz, S., And Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2. John Wiley & Sons, New York.

Johnson, R.A. (1970): Asymptotic Expansions Associated With Posterior Distributions. Annals Of Mathematical Statistics, 41, 196-215.

Kalbfleisch, J.D., And Prentice, R.L. (2002). The Statistical Analysis Of Failure Time Data. John Wiley & Sons, New York.

Khan,A.A., Puri, P.D., and Yaquab,M.(1996): Approximate Bayesian inference in location – scale models. Proceedings of national seminar on Bayesian statistics and applications, April6-8,1996,89-101,Dept of statistics BHU, Varanasi.

Khan, A. A.(1997). Asymptotic Bayesian analysis in location-scale models. Ph.D. Thesis submitted to the department of mathematics and statistics, HAU, Hisar.

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).

Laplace, P. S. (1986). Memoir on the probability of causes of events. Statistical science 1, 364-378 (translation by S. Stigler of Memoire Sur La Probabilities Des Causes Parles Evenemens).

Laplace,P.S, (1812): Theorie Analytiquie des probabilities, paris: courair.

Lawless, J.F. (2003). Statistical Models And Methods For Lifetime Data. Wiley, New York.

Lee, P. M. (1997). Bayesian Statistics An Introduction. John Wiley & Sons Inc. New York. Toronto, 50-53,278-279.

Lindley, D. V. (1958). Fiducial distributions and Bayes theorem. Journal of the royal statistical society, B, 20, 102-107.

Lindely, D. V. (1971): Bayesian Statistics, A Review. Siam, Philadelphia.

Lindley, D.V. (1961): “The Use Of Prior Probability Distribution In Statistical Inference & Decision”,In The Proceeding Of The Fourth Berkeley Symposium (Vol-I), Berkeley: University Of California Press, Pp. 453-468.

Lindley, D.V. (1965): Introduction to probability and statistics fom a Bayesian view point(partI and II) Cambridge University press.

Lindley,D.V.(1980): “Approximate Bayesian Method”. Traabayos Estatistica 31, 223-237.

Mendenhall, W. (1958). A Bibliography On Life Testing And Related Topics. Biometrika 45, 521-543.

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.

Morvie, A. De.(1738): Approximatio Ad Summan Ferminorum Binomii $(a+b)^n$ In Seriem Expansi, Supplementum Ii To Miscellanae Analytica, 1-7.

Mosteller, F. And Wallace, D.L. (1964): Applied Bayesian And Classical Inference. Springer-Verlag, New York.

Pairman,E. (1919): Tables Of The Di-Gamma And Tri-Gamma Functions In Tracts For Computers No.1, Karl Pearson (Ed), Cambridge University Press.

Pratt, J. W. (1965). Bayesian interpretation of standard inference statements (with discussion). Journal of the royal statistical society B 27, 169-203.

Raffia, H., and Schlaifer, R. (1961). Applied statistical decision theory, Boston, Massachusetts: Harvard Business School.

Rubin, D.B., And Schenker, N. (1987). Logit-Based Interval Estimation For Binomial Data Using The Jeffrey’s Prior. Sociological Methodologies, 131-144.

Saal,N.Z.M., A.A.Jemain and S.H.A.Al-Mashoor,(2008): A Copmparision of Weibull and Gamma Distribution in Application of Sleep Apnea Asia J.Math,Statist., 1: 132-138.

Sinha, S.K. and Kale,B.K.(1980): Life Testing And Relaibility Estimation.Wiley Eastern/Halsted Press,Delhi/New York.

Sinha, S.K. and Howlader.H.A.(1983): Credible And HPD Intervals of the Parameters and Relaibility of Ray leigh Distribution. IFEE Transaction Relaibility R-32,2,217-220.

Sinha, S.K. (1986). Reliability And Life Testing. Wiley Eastern Limited.

Sinha, S.K. (1998). Bayesian Estimation . New Age International (P) Limited, Publishers, New Dehli. Somepith And Of Arbitrary Phase. Philosophical Magazine, 5-Th Series, 10,73-78.Sons.

Stigler, S. (1986): The History Of Statistics, Beknap, Harvard.

Sukhatmi, P.V.(1937): Tests Of Significance For Samples Of The x^2-Population With 2 Degrees Of Freedom, Annals Of Eugenice, London, 8,52-56.

Tierney, L., And Kadane, J.B. (1986). Accurate Approximations For Posterior Moments And Marginal Densities. J. Amer. Statist. Assoc., 81, 82-86.

Tierney, L., Kass, R.E.,And Kadane,J.B.(1989a): Approximate Marginal Densities of non –linear Functions", biometrika,76,425-433.

Wong, W.A And B. Li.(1992): Laplace Expansion For Posterior Densities Of Non-Linear Function Of Parameters Biometria, 79,393-398.

Zaman, M.R., M.K.Roy and N.Akhter.(2005): Chi-Square Mixture of Gamma Distribution,J.Applied Sci, 5: 1632-1635.

Zellener, A (1971). An introduction to Bayesian inference in econometrics. Wiley, New York.