A Study on Bayesian Estimation of Parameters of Some Well Known Distribution Functions

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

by

Ms. Subhasmita Sahoo

Under the guidance of

Prof. Manas Ranjan Tripathy

Department of Mathematics
National Institute of Technology
Rourkela-769008
India

May 2014

Certificate

This is to certify that the thesis entitled **"A Study on Bayesian Estimation of Parameters of Some Well Known Distribution Functions"**, which is being submitted by **Subhasmita Sahoo** in the Department of Mathematics, National Institute of Technology, Rourkela, in partial fulfilment for the award of the degree of **Master of Science**, is a record of bonafide review work carried out by her in the Department of Mathematics under my guidance. She has worked as a project student in this Institute for one year. In my opinion the work has reached the standard, fulfilling the requirements of the regulations related to the Master of Science degree. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Place: NIT Rourkela
Date: May 2014

Dr. M. R. Tripathy
Assistant Professor
Department of Mathematics
NIT Rourkela-769008
India

Acknowledgement

It is a great pleasure and proud privilege to express my deep sense of gratitude to my guide, Prof. M. R. Tripathy. I am grateful to him for his, continuous encouragement and guidance throughout the period of my project work. Without his active guidance it would not have been possible for me to complete the work.

I acknowledge my deep sense of gratitude to the Head of the Department, all faculty members, all nonteaching staff members of the department of mathematics and the authorities of NIT Rourkela.
I also thank to all my friends for their support, co-operation and sincere help in various ways. I express my sincere thank with gratitude to my parents and other family members, for their support, blessings and encouragement. Finally, I bow down before the almighty who has made everything possible.

Place: NIT Rourkela
Date: May 2014

(Subhasmita Sahoo)
Roll No-412MA2080

Abstract

The thesis addresses the problem of estimation of parameters of some well known distribution functions. The problem of estimation of parameters of binomial, Poisson, normal and exponential distribution function has been considered. In particular, the maximum likelihood, method of moment, and Bayes estimators has been derived. Further the problem of estimating the parameter of binomial, Poisson, normal and exponential distribution function by Lindley's Approximation is considered. Similar type of estimators are also derived for this case using different types of prior.

Contents

Chapter 1

Introduction and Motivation

The problem of estimation of parameters of a distribution function is one of the important and major areas in the area of statistical inference. Statistical inference is the process by which information from a sample data is used to draw conclusions about the population from which the sample was selected. The theory of point estimation was first studied by prof. R. A. Fisher. Estimation theory traces its origin to the efforts of astronomers many years ago to predict the motion of our solar system. Estimate is calculated approximately of the result which is given even if the input data is uncertain. It is one of the core topics in mathematical statistics. This problem of estimation finds applications in industries, stock markets, business analytics, social sciences, socioeconomic study, reliability study etc. In various agricultural, physical and industrial experiments, one comes across situations, where parameters associated with the data are to be estimated. For example,

(i) In business, a chamber of commerce may want to know the average income of the families in its community.

(ii) In science, a mineralogist may wish to determine the average iron content of a given core.

(iii) Suppose a manufacturer of tube lights wants to know the average life times of certain bulbs. In this case the data which will be taken randomly may follow an exponential distribution, and hence we may estimate the average life span of the tube lights.

Basically two types of estimation procedures are known. One is the point estimation and another is interval estimation or confidence interval. Here we mainly focus on point estimation of the parameters associated with a distribution function. This normally refers to the process of approximating a parameter (which is assumed to be unknown) using the sample data which may have certain probability distribution. Suppose some observed data X follows $N(\mu, \sigma^2)$, where μ is unknown and σ^2. let us take a sample $X_1, X_2, ..., X_n$ from $X \sim N(\mu, \sigma^2)$, where n is the sample size. The statistic $T = \sum_{i=1}^{n} X_i$ is the best estimate for μ. For sufficient statistic please see Chapter 2. The value of $T(X_1, X_2, \ldots, X_n)$ for given sample values $x_1, x_2, ..., x_n$ is the estimate of μ. The process of estimating the parameters by using the sample values is known as estimation. Suppose we are interested to know the quality of production of rice across the country (say India) in last ten years. If the collected data follows normal distribution, then by estimating the parameter μ we can have an idea about the average rice production during that period and and estimating the parameter σ^2 we can talk about the variability of production of rice in the country.

Sometimes we may get a better approximate regarding the parameter by having some prior information about the parameter. This type of study is known as Bayesian study. In this process we assume that the parameter has certain distribution that is the parameter which has to be estimated is considered as a random variable. For this purpose we may consider informative prior or non-informative prior for the unknown parameter. For example, the $Binomial(n, p)$ distribution taking prior as $g(p) = 1;\ 0 < p < 1$. It is a noninformative prior. In this project work we have discussed some discrete distribution function and estimation of its parameters. The rest of the work can be organized in the following way.

The main goal of this project work is to learn different estimation techniques for estimating the parameters. In Chapter 2, we have discussed some basic results related to the point estimation and Bayesian estimation which will be usefull in the subsequent chapters. We study the problem of estimating the parameters of a binomial, Poisson, normal and exponential in Chapter 3. In Chapter 4, we learn a different technique for obtaining Bayes estimator. The technique is due to Lindley (1980) which is known for approximating a ratio of integrals.

Chapter 2

Definitions and Basic Results

In this chapter some definitions and basic results are discussed which are very much essential for the development of the entire project work. Below we start from a very basic concept known as random experiment or statistical experiment which arises in either nature or by some statistician.

2.1 Some Basic Definitions

Definition 2.1 (Random experiment) *An experiment in which all outcomes are known in advance, any performance of the experiment that results in an outcome is not known in advance and the experiment can be repeated under identical conditions, is called a random experiment.*

Definition 2.2 (Sample space) *The sample space of a statistical experiment is a pair* (Ω, S)*, where* Ω *is the set of all possible outcomes of an experiment and* S *is the* σ*-field of subsets of* Ω*.*

Definition 2.3 (Event) *An event is a subset of the sample space* Ω *in which we are interested. Any set* $A \in S$ *is known as the events.*

Definition 2.4 (Probability measure) *Let (Ω, S) be a sample space and S be the $\sigma-$ algebra defined over Ω. A set function P defined on S is called probability measure or simply probability if it satisfies the following conditions,*

(i) $P(A) \geq 0, \forall\ A \in S.$

(ii)$P(\Omega) = 1.$

(iii) Let A_j , $A_j \in S, j = 1, 2, \ldots$ be a disjoint sequence of sets. That is $A_j \bigcap A_k = \emptyset$ for $j \neq k$. Then

$$P\Big(\bigcup_{j=1}^{\infty}\Big) = \sum_{j=1}^{\infty} P\Big(A_j\Big).$$

Definition 2.5 (Random variable) *Let (Ω, S) be a sample space. A finite single valued function which maps Ω into $\mathbb{R}$ is called a random variable if the inverse images under X of all Borel sets in $\mathbb{R}$ are events.*

Definition 2.6 (Distribution function) *Let X be a random variable defined on (Ω, S, P). Define a function F on $(-\infty, \infty)$ by $F(x) = P\{w : X(w) \leq x\}$ for all $x \in \mathbb{R}$. F is nondecreasing, right continues and $F(-\infty) = 0$, $F(\infty) = 1$. Then the function F is called the distribution function of the random variable X.*

Depending upon the nature of sample space we may categorize the random variables as discrete type and continuous type.

Definition 2.7 (Discrete random variable) *An rv X defined on (Ω,S, P) is said to be of discrete type or simply discrete if there exist a countable set $E \subseteq \mathbb{R}$ such that $P(x \in E) = 1$. The collection of numbers $\{p_i\}$ satisfying $P\{X = x_i\} = p_i \geq 0$ for all i and $\sum_{i=1}^{\infty} p_i = 1$, is called the probability mass function of an rv X.*

Definition 2.8 (Continuous random variable) *Let X be a random variable defined on (Ω,S, P) with distribution function F. Then X is said to be continuous random variable* *if F is absolutely continuous that is if there exists a nonnegative function $f(x)$ such that, for every real number x, we have $F(x) = \int_{-\infty}^{x} f(t)\,dt$. The function f is called the probability density function of the random variable X and $\int_{-\infty}^{\infty} f(t)\,dt = 1$.*

Next we discuss some basic results and terminologies related to the estimation of parameters of a distribution function.

2.2 Estimation of Parameters of a Distribution Function

In this project we will only concentrate on the problem of point estimation on a classical point of view and some of its extensions to Baysian studies.

Suppose $X_1, X_2, \ldots, X_n$ are collected from a population which has a distribution function $F_\theta(x)$. This is normally a family of distribution functions as for each value of the parameter θ we have a distribution F. Here θ may lie in some set say Θ (parameter space), is the unknown parameter associated with the distribution function F. Our aim is to get an approximate value or an estimate of θ using the samples X. We will study the theory of point estimation (classical approach) and particularly parametric point estimation.

Definition 2.9 (Parameter space) *The set of all possible values of the parameters of a distribution function F is called the parameter space. This set is usually denoted by Θ.*

Definition 2.10 (Statistic) *Any function of the random sample $X_1, X_2, \ldots, X_n$ those are being observed say $T(X_1, X_2, \ldots, X_n)$ is called a statistic. The value of a statistic is denoted by $T(x_1, x_2, \ldots, x_n)$.*

Definition 2.11 (Estimator) *If a statistic is used to estimate an unknown parameter θ of a distribution, then it is called an estimator and a particular value of the estimator say $T_n(X_1, X_2, \ldots, X_n)$ is called an estimate of θ.*

Next, we discuss some of the properties which an estimator may enjoy.

2.3 Characteristics of Estimators

Various statistical properties of an estimators can be used to decide which estimator is most appropriate in a given situation.

Definition 2.12 (Unbiasedness) *: A statistic $T(X_1, X_2, \ldots, X_n)$ is an unbiased estimator of the parameter θ if and only if $E[T(X_1, X_2, \ldots, X_n)] = \theta$. If $E[T(X_1, X_2, \ldots, X_n)] > \theta$ then we say T over estimates and if $E[T(X_1, X_2, \ldots, X_n)] < \theta$ we say it underestimates.*

Example 2.1 *Let X_1, X_2, X_3 be a random sample of size 3 from a normal population $N(\mu, \sigma^2)$. Suppose μ, and σ^2 are unknown. It can be easily seen that a statistic $T = \frac{1}{4}(X_1 + 2X_2 + X_3)$ is an unbiased estimate of μ. Since,*

$$\begin{aligned} E(T) &= E\left(\left[\frac{1}{4}(X_1 + 2X_2 + X_3\right]\right) \\ &= \frac{1}{4}(\mu + 2\mu + \mu) \\ &= \mu. \end{aligned}$$

Definition 2.13 (Consistency) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with common distribution function F_θ ,$\theta \in \Theta$. A sequence of point estimators $T_n(X_1, X_2, \ldots, X_n) = T_n$ will be called consistent for $\psi(\theta)$ if T_n converges to $\psi(\theta)$ in probability, that is,*

$$P(|T_n - \psi(\theta)| > \epsilon) \to 0, \text{ as } n \to \infty,$$

where ϵ is a very small arbitrary positive number.

Below we discuss some of the consequences of the above definition.

Remark 2.1 *If T_n is a consistent estimator of θ and $\psi(\theta)$ is a continuous function of θ, then $\psi(T_n)$ is a consistent estimator of $\psi(\theta)$.*

Remark 2.2 *If T_n is a sequence of consistent estimators such that $E[T_n] \to \psi(\theta)$ and $Var[T_n] \to 0$ as $n \to \infty$, then T_n is a consistent estimator of $\psi(\theta)$.*

Next we discuss a property of an estimator which tells us how much an estimator is good with respect to another estimator.

Definition 2.14 (Efficiency) *In general it is possible to have more than one consistent estimators among all unbiased estimators. Thus it is necessary to find some criteria to choose between the estimators. Such a criterion which is based on the variances of sampling distributions of estimators is known as efficiency.*

Definition 2.15 (Sufficiency) *An estimator is said to be sufficient for a parameter θ, if it contains all the information in the sample regarding the parameter. Let $X = (X_1, X_2, \ldots, X_n)$ be a sample from a family of distributions $F_\theta : \theta \in \Theta$. A statistic T is sufficient for θ if and only if the conditional distribution of X given $T = t$, does not depend upon θ.*

Next, we discuss a technique known as "Fisher-Neymann factorization criterion" to determine sufficient statistics for a given distribution.

Theorem 2.1 (Fisher-Neymann Factorization Criterion) *A statistic $T = t(X)$ is a sufficient statistic for the parameter θ if and only if the joint probability distribution or density of the random sample can be expressed in the form:*

$$f(x_1, x_2, \ldots, x_n; \theta) = g_\theta(t(x)) \times h(x_1, x_2, \ldots, x_n),$$

where $g_\theta(t(x))$ depends on θ and x and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ.

2.4 Methods of Estimation

Normally there are two different approaches for obtaining a point estimator for unknown parameter θ. Namely classical method and decision theoretic approach. Now we outline some of the most important methods for obtaining point estimators. Most commonly used methods under classical estimation are as follows.

2.4.1 Method of Moments

Suppose X be a random variable with distribution function F. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from X. Defining the first k sample moments about origin as $m_r^{'} = \frac{1}{n}\sum_{i=1}^{n} X_i^r$, $r = 1, 2, \ldots, k$. The first k population moments about origin are given by $\mu_r^{'} = E(X^r)$, which are in general functions of k unknown parameters. Equating the sample moments and population moments yields k simultaneous equations in k unknowns. $\mu_r^{'} = m_r^{'}$, $r = 1, 2, \ldots, k$. The solutions to the above equations denoted by $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ yields the moment estimators of $\theta_1, \theta_2, \ldots, \theta_k$.

Example 2.2 *Extimate α and β in the case of Pearson's Type III distribution by the method of moments:*

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},\ 0 \leq x < \infty.$$

Solution: *We have*

$$\begin{aligned}
\mu_r^{'} &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^r x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{r+\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+r)}{\beta^{\alpha+r}} \\
&= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)\beta^r}
\end{aligned}$$

$$\begin{aligned}
\mu_1^{'} &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} \\
&= \frac{\alpha}{\beta'} \\
\mu_2^{'} &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\beta^2} \\
&= \frac{(\alpha+1)\alpha}{\beta^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\mu_2^{'}}{\mu_1^{'2}} &= \frac{\alpha+1}{\alpha} \\
&= \frac{1}{\alpha}+1.
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \frac{1}{\alpha} &= \frac{\mu_2^{'}}{\mu_1^{'2}-1} \\
\Rightarrow \frac{1}{\alpha} &= \frac{\mu_2^{'}-\mu_1^{'2}}{\mu_1^{'2}} \\
\Rightarrow \alpha &= \frac{\mu_1^{'2}}{\mu_2^{'}-\mu_1^{'2}}
\end{aligned}$$

We know that

$$\begin{aligned}
\frac{\alpha}{\beta} &= \mu_1^{'} \\
\Rightarrow \beta &= \frac{\alpha}{\mu_1^{'}} \\
\Rightarrow \beta &= \frac{\frac{\mu_1^{'2}}{\mu_2^{'}-\mu_1^{'2}}}{\mu_1^{'}} \\
\Rightarrow \beta &= \frac{\mu_1^{'}}{\mu_2^{'}-\mu_1^{'2}}.
\end{aligned}$$

Hence $\hat{\alpha} = \frac{m_1^{'2}}{m_2^{'}-m_1^{'2}}$ *and* $\hat{\beta} = \frac{m_1^{'}}{m_2^{'}-m_1^{'2}}$
where $m_1^{'}$ *and* $m_2^{'}$ *are sample moments.*

2.4.2 Method of Maximum Likelihood Estimation

Suppose $(X_1, X_2, \ldots, X_n)$ be a random vector with PDF $f_{\underline{\theta}}(x_1, x_2, \ldots, x_n)$, $\underline{\theta} \in \Theta$, where $\underline{\theta}$ is a multidimensional vector valued unknown parameter. The likelihood function is given by $L(\underline{\theta}; x_1, x_2, \ldots, x_n) = f_{\underline{\theta}}(x_1, x_2, \ldots, x_n)$ which is a function of unknown parameter $\underline{\theta}$ for fixed sample sizes. If $X_1, X_2, \ldots, X_n$ are iid with PDF $f_{\underline{\theta}}(x)$, then the likelihood function is

$$L(\underline{\theta}; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{\underline{\theta}}(x_i).$$

The maximum likelihood estimator (MLE) of $\underline{\theta}$ is the value of $\underline{\theta}$ say $\hat{\theta}$ that maximizes the likelihood function $L(\underline{\theta}; x_1, x_2, \ldots, x_n)$. Note that in many cases, the likelihood function can be infinitesimal and it is much easier to deal with the log-likelihood function that is $\log L(\underline{\theta}; x_1, x_2, \ldots, x_n)$. Since log is a monotone function, when likelihood function is maximized, log-likelihood function is also maximized, and vice versa.

Example 2.3 *In random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimators for*
(i) μ when σ^2 is known,
(ii) σ^2 when μ is known.

Solution: *It is given that $X \sim N(\mu, \sigma^2)$, then the likelihood function is,*

$$\begin{aligned} L(\mu, \sigma^2 | (\underline{x})) &= \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x_i - \mu)^2 \right\} \right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}. \end{aligned}$$

Taking logarithm of the likelihood function we have,

$$\begin{aligned} \log L &= \log\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n + \log\left[\exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \right] \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2. \end{aligned}$$

(i) When σ^2 is known, the likelihood equation for estimating μ is:

$$\frac{\partial}{\partial\mu}\log L = 0$$
$$\Rightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$
$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0$$
$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

(ii) When μ is known, the likelihood equation for estimating σ^2 is:

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

Remark 2.3 *Let T be a sufficient statistic for the family of probability density functions $f_\theta(x)$; $\theta \in \Theta$. If an MLE of θ exists, it is a function of T.*

Remark 2.4 *If MLE exists then it is the most efficient in the class of such estimators.*

Remark 2.5 (Invariance property) *If T is the MLE of θ and $\psi(\theta)$ is one-to-one function of θ, then $\psi(T)$ is the MLE of $\psi(\theta)$.*

When we estimate the unknown parameter θ of a distribution function $F_\theta(x)$, by an estimator $\delta(X)$ some loss is incurred. Hence we use some loss functions to know the amount of loss incurred as below.

Definition 2.16 (Loss Function) *Loss function represents the loss incurred when the true value of the parameter is θ and we are estimating θ by $\delta(x)$. Throughout the discussion the loss function $L(\theta, \delta(x))$ is taken as nonnegative and real valued in both its arguments. When the correct estimate is chosen the loss becomes zero. Depending on the loss function Bayes estimators are different. Different types of loss functions are discussed below.*

Definition 2.17 (Linear Loss Function) *The linear loss function is defined as*

$$\begin{aligned} L(\theta, \delta(x)) &= c_1\Big(\delta(x) - \theta\Big), \delta(x) \geq \theta \\ &= c_2\Big(\theta - \delta(x)\Big), \delta(x) < \theta \end{aligned}$$

The constants c_1 and c_2 reflect the effect over and under estimating θ. If c_1 and c_2 are functions of θ, the above loss function is called weighted linear loss function.

Definition 2.18 (Absolute Error Loss Function) *The absolute error loss function is defined as*

$$L(\theta, \delta(x)) = |\delta(x) - \theta|.$$

Definition 2.19 (Squared Error Loss Function) *The squared error loss function is defined as*

$$L(\theta, \delta(x)) = k(\delta(x) - \theta)^2.$$

It is also called as quadratic loss function.

Throughout our discussion we have used squared error loss function.

Definition 2.20 (Risk Function) *The average loss of an estimator $\delta(x)$ is known as its risk function and is defined as*

$$R(\theta, \delta) = E[L(\theta, \delta(x))].$$

The goal of an estimation problem is to look for an estimator δ which has uniformly minimum risk for all values of the parameter $\theta \in \Theta$.

2.4.3 Bayesian Estimation

In Bayesian Principle the unknown parameter θ which is treated as random variable assumes a probability distribution known as a priori of θ denoted by $\Pi(\theta)$.

To start the estimation of parameters we have the prior information about the unknown parameter θ. Different types of prior are discussed below.

(a) Noninformative Prior: A probability density function $\Pi(\theta)$ of θ is said to be a noninformative prior if it contains no information about θ. Some simple examples of noninformative priors are $\Pi(\theta) = 1$, and $\Pi(\theta) = \frac{1}{\theta}$.

(b) Natural conjugate prior: To simplify the calculations, statisticians use natural conjugate priors. Usually there is a natural parameter family of distributions such that the posterior distributions also belong to the same family. These priors make the computations much simpler. Conjugate priors are usually associated with the exponential family of distributions. Some example of natural conjugate priors are: with sampling from pdf $N(\theta, \sigma^2)$ we take prior distribution $N(\mu, \tau^2)$, the posterior distribution is

$$N\Big(\frac{\sigma^2\mu + x\tau^2}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\Big).$$

With sampling distribution Binomial and prior distribution Beta the posterior distribution is Beta. For some more results one may see the Book by Rohatgi and Saleh (2003).

(c) Jeffreys' invariant prior: Sir Harold Jeffreys suggested a general rule for choosing the non-informative prior for θ,

$$\Pi(\underline{\theta}) \propto \sqrt{I(\underline{\theta})},$$

where $\underline{\theta}$ vector valued parameter, and

$$I(\underline{\theta}) = -E\Big[\frac{\partial^2 \log f(x|\underline{\theta})}{\partial\theta_i \partial\theta_j}\Big],$$

where $I(\underline{\theta})$ is Fisher information matrix.

Definition 2.21 (Posterior distribution) *The posterior distribution of θ given $X = x$ is obtained by dividing the joint density of θ and X by the marginal distribution of X. Mathematically*

$$\frac{\Pi(\theta) f_\theta(x)}{\int_\Theta \Pi(\theta) f_\theta(x) d\theta}$$

where Θ is the parameter space.

Definition 2.22 (Bayes risk) *Bayes risk associated with an estimator δ is defined as the expected value of the risk function $R(\theta, \delta)$ with respect to the prior distribution $\Pi(\theta)$ of θ and is given by,*

$$\begin{aligned} R^*(\theta, \delta) &= E[R(\theta, \delta)] \\ &= \int R(\theta, \delta)\Pi(\theta) d\theta \\ &= \int E[L(\theta, \delta)]\Pi(\theta) d\theta. \end{aligned}$$

Definition 2.23 (Bayes estimator) *A Bayes estimator is that which minimizes the Bayes risk defined above. Accordingly if δ_o is Bayes estimator of θ with prior distribution $\Pi(\theta)$, then we must have*

$$R^*(\theta, \delta_o) = \inf R^*(\theta, \delta).$$

Theorem 2.2 *The Bayes estimator of a parameter $\theta \in \Theta$ with respect to the qudratic loss function $L(\theta, \delta) = (\theta - \delta)^2$ turns out to be*

$$\delta(x) = E\{\theta | X = x\}.$$

Chapter 3

Bayesian Estimation of Parameters of Binomial, Poisson, Normal and Exponential Distribution

In this chapter, we consider the problem of estimating the parameters of some well known distributions such as binomial, Poisson, Exponential and Normal. We consider some non-informative priors and conjugate priors for their parameters. We first take the binomial distribution and study the estimation of their parameters.

3.1 Binomial Distribution

Let $X_1, X_2, \ldots, X_N$ be N samples taken from binomial distribution with parameter n and p. The method of moment estimator of the parameter p is given by $\frac{1}{n}\sum_{i=1}^{N} x_i$. This is also the MLE of p. Now we try to get a Bayes estimator of parameter p.

The probability mass function of the random variable X is given by,

$$f(x|(n,p)) = \binom{n}{x} p^x q^{n-x},\ \ q = 1-p;\ \ x = 0, 1, 2, ..., n.$$

The likelihood function is given by,

$$\begin{aligned} L((x_1, x_2, \ldots, x_N), p) &= \prod_{i=1}^{N} \binom{n}{x_i} p^{x_i} q^{n-x_i} \\ &= p^s (1-p)^{Nn-S} \prod_{i=1}^{N} \binom{n}{x_i}, \end{aligned}$$

where $S = \sum_{i=1}^{N} x_i$.
Consider the conjugate prior distribution of p :

$$g(p) \propto p^{a-1}(1-p)^{b-1}, \ a, \ b > 0.$$

The joint probability distribution of p and X is given by,

$$h(x, p) = K p^{S+a-1}(1-p)^{Nn+b-S-1} \prod_{i=1}^{N} \binom{n}{x_i}.$$

The marginal probability distribution is given by,

$$\begin{aligned} f_X(x) &= \int_0^1 h(x, p) dp \\ &= K \prod_{i=1}^{N} \binom{n}{x_i} \int_0^1 p^{S+a-1}(1-p)^{Nn+b-S-1} dp \\ &= K \prod_{i=1}^{N} \binom{n}{x_i} B(S+a, Nn+b-S). \end{aligned}$$

Now the posterior probability is given by,

$$\Pi(p|\underline{x}) = \frac{1}{B(S+a, Nn+b-S)} p^{S+a-1}(1-p)^{Nn+b-S-1}.$$

Hence Bayes estimator of p is given by,

$$\begin{aligned}
p^* &= E(p|x^*) \\
&= \int_0^1 p\Pi(p|\underline{x})dp \\
&= \int_0^1 p\frac{1}{B(S+a, Nn+b-S)} p^{S+a-1}(1-p)^{Nn+b-S-1}dp \\
&= \frac{1}{B(S+a, Nn+b-S)} \int_0^1 p^{S+a}(1-p)^{Nn+b-S-1}dp \\
&= \frac{B(S+a+1, Nn+b-S)}{B(S+a, Nn+b-S)} \\
&= \frac{\Gamma(S+a+1)\Gamma(Nn+b-S)}{\Gamma(S+a+Nn+b-S+1)} \frac{\Gamma(S+a+Nn+b-S)}{\Gamma(S+a)\Gamma(Nn+b-s)} \\
&= \frac{(S+a)!(S+a+Nn+b-S-1)!}{(S+a-1)!(S+a+Nn+b-S)!} \\
&= \frac{S+a}{a+b+Nn}.
\end{aligned}$$

3.2 Poisson Distribution

Let $x = (x_1, x_2, ..., x_n)$ be a random sample from Poisson distribution. The method of moment and the MLE of parameter λ is $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Now we try to get a Baye estimator of paramere λ by the following way.

The probability mass function of the random variable is given by,

$$f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, 2, ...$$

The likelihood function,

$$L(\lambda|\underline{x}) = \frac{e^{-n\lambda}\lambda^S}{x_1!x_2!...x_n!}, \ S = \sum_{i=1}^{n} x_i.$$

Consider the natural conjugate prior(NCP) for λ is given by,

$$g(\lambda|a, b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}, \ a, b, \lambda > 0.$$

The joint pdf is given by,

$$h(x,\lambda) = \frac{b^a}{\Gamma(a)x_1!, x_2!, ..., x_n!}\lambda^{S+a-1}e^{-\lambda(n+b)}.$$

The marginal pdf of λ is given by,

$$\begin{aligned}
f_X(x) &= \int_0^\infty h(x,\lambda)d\lambda \\
&= \frac{b^a}{\Gamma(a)x_1!, x_2!, ..., x_n!}\int_0^\infty \lambda^{S+a-1}e^{-\lambda(n+b)}d\lambda \\
&= \frac{b^a}{\Gamma(a)x_1!, x_2!, ..., x_n!}\int_0^\infty \left(\frac{t}{n+b}\right)^{S+a-1} e^{-t}\frac{dt}{n+b}, \quad (put \quad t=(n+b)\lambda) \\
&= \frac{b^a}{\Gamma(a)x_1!, x_2!, ..., x_n!(n+b)^{S+a}}\int_0^\infty t^{S+a-1}e^{-t}dt \\
&= \frac{b^a\Gamma(S+a)}{\Gamma(a)x_1!, x_2!, ..., x_n!(n+b)^{S+a}}.
\end{aligned}$$

Now the posterior distribution is given by,

$$\Pi(\lambda|\underline{x}) = \frac{(n+b)^{S+a}}{\Gamma(S+a)}\lambda^{S+a-1}e^{-\lambda(n+b)}.$$

The Bayes estimator of λ with respect to the squared error loss function is given by,

$$\begin{aligned}
\lambda^* &= E(\lambda|\underline{x}) \\
&= \frac{(n+b)^{S+a}}{\Gamma(S+a)}\int_0^\infty \lambda.\lambda^{S+a-1}e^{-\lambda(n+b)}d\lambda \\
&= \frac{(n+b)^{S+a}}{\Gamma(S+a)}\int_0^\infty \left(\frac{t}{n+b}\right)^{S+a} e^{-t}\frac{dt}{n+b}, \quad (put \quad t=(n+b)\lambda) \\
&= \frac{(n+b)^{S+a}}{\Gamma(S+a)(n+b)^{S+a+1}}\int_0^\infty t^{(S+a+1)-1}e^{-t}dt \\
&= \frac{(n+b)^{S+a}\Gamma(S+a+1)}{\Gamma(S+a)(n+b)^{S+a+1}} \\
&= \frac{(n+b)^{S+a}(S+a)!}{(n+b)^{S+a+1}(S+a-1)!} \\
&= \frac{S+a}{n+b}.
\end{aligned}$$

3.3 Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed random samples taken from normal distribution $X \sim N(\mu, \sigma^2)$. Here the method of moments of estimator μ and σ^2 are $\hat{\mu} = \bar{X}$ and $\hat{\sigma^2} = \frac{S^2}{n}$ which is also the MLE of μ and σ^2. Now we try to get the Bayes estimator of parameters μ and σ^2.

The probability density function of the random variable X is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty; \ \sigma > 0; \ -\infty < \mu < \infty.$$

Taking $\sigma^2 = 1$ we estimate the unknown parameter μ. The likelihood function is given by,

$$\begin{aligned} L(\underline{x}, \mu) &= \prod_{i=1}^{n} \Big[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} \Big] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\frac{-1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + \mu^2 \right)}. \end{aligned}$$

Let the prior PDF of μ be $N(0,1)$ which is given by,

$$g(\mu) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\mu^2}{2}}.$$

The joint PDF of $\underline{X}$ and μ is given by,

$$f(\underline{x}, \mu) = \frac{1}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-(\sum_{i=1}^{n} x_i^2 - 2\mu n\bar{x} + (n+1)\mu^2)}{2}}.$$

The marginal PDF of $\underline{X}$ is given by,

$$\begin{aligned} h(\underline{x}) &= \int_{-\infty}^{\infty} f(x, \mu) d\mu \\ &= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-(\sum_{i=1}^{n} x_i^2 - 2\mu n\bar{x} + (n+1)\mu^2)}{2}} d\mu \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-1}{2}\sum_{i=1}^{n} x_i^2} \int_{-\infty}^{\infty} e^{\frac{-(n+1)}{2}[(\mu - \frac{n\bar{x}}{n+1})^2 - (\frac{n\bar{x}}{n+1})^2]} d\mu \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-1}{2}\sum_{i=1}^{n} x_i^2} e^{(\frac{n}{2})(\frac{n\bar{x}}{n+1})^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(n+1)}{2}(\mu - \frac{n\bar{x}}{n+1})^2} d\mu \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-1}{2}(\sum_{i=1}^{n} x_i^2 - \frac{n^2\bar{x}^2}{n+1})} \frac{1}{(n+1)^{\frac{1}{2}}}. \end{aligned}$$

Now the posterior PDF is given by,

$$\begin{aligned} f(\mu|\underline{x}) &= \frac{f(\underline{x},\mu)}{h(\underline{x})} \\ &= \frac{(n+1)^{\frac{1}{2}}}{\sqrt{2\pi}} e^{\frac{-(n+1)}{2}(\mu-\frac{n\bar{X}}{n+1}))^2}. \end{aligned}$$

The Bayes estimator of μ with respect to the squared error loss function $L(\mu,\delta) = (\mu-\delta)^2$, is given by,

$$\begin{aligned} \hat{\mu} &= E(\mu|\underline{x}) \\ &= \int_{-\infty}^{\infty} \mu f(\mu|\underline{x}) d\mu \\ &= \frac{(n+1)^{\frac{1}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{-(\frac{n+1}{2})(\mu-\frac{n\bar{X}}{n+1})^2} \\ &= \frac{n\bar{X}}{n+1}. \end{aligned}$$

3.4 Exponential Distribution(one-parameter)

In this section we will discuss the estimation problem for one-parameter exponential distribution.

Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed random samples taken from one-parameter exponential distribution $Ex(\beta)$. The method of moments of estimator and MLE of the parameter β is $\bar{X}$. Now we try to get the Bayes estimator of the parameter β as following.
The probability density function of the random variable X is given by,

$$f(x;\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \quad x > 0, \quad \beta > 0.$$

The likelihood function is given by,

$$
\begin{aligned}
L(\underline{x};\beta) &= \prod_{i=1}^{n} \frac{e^{\frac{-x_i}{\beta}}}{\beta} \\
&= \frac{1}{\beta^n} e^{\frac{-1}{\beta}\sum_{i=1}^{n} x_i} \\
&= \frac{1}{\beta^n} e^{\frac{-S}{\beta}}\ , \quad \text{where} \ \ S = \sum_{i=1}^{n} x_i.
\end{aligned}
$$

Considering the inverted gamma prior, the prior pdf of β is given by,

$$
g(\beta|a,b) = \frac{a^b}{\Gamma(b)} \frac{e^{\frac{-a}{\beta}}}{\beta^{b+1}},\ \ a > 0,\ \ b > 0,\ \ \beta > 0.
$$

The joint pdf of X and β is given by,

$$
f(\underline{x},\beta) = \frac{a^b}{\Gamma(b)} \frac{e^{\frac{-1}{\beta}(S+a)}}{\beta^{n+b+1}},\ \ a > 0,\ \ b > 0,\ \ \beta\ \ > 0.
$$

The marginal pdf of X is given by,

$$
\begin{aligned}
h(\underline{x}) &= \int_0^\infty f(\underline{x},\beta)d\beta \\
&= \frac{a^b}{\Gamma(b)} \int_0^\infty \frac{e^{\frac{-1}{\beta}(S+a)}}{\beta^{n+b+1}} d\beta, \\
&= \frac{a^b}{\Gamma(b)} \int_0^\infty e^{-t} \frac{t^{n+b+1}}{(S+a)^{n+b}} dt, \quad (\text{put} \ \ t = \frac{S+a}{\beta}), \\
&= \frac{a^b \Gamma(n+b)}{\Gamma(b)(S+a)^{n+b}}.
\end{aligned}
$$

Now the posterior PDF is given by,

$$
f(\beta|\underline{x}) = \frac{(S+a)^{n+b} e^{\frac{-(S+a)}{\beta}}}{\beta^{n+b+1}\Gamma(n+b)}.
$$

The Bayes estimator of β with respect to the squared error loss function $L(\beta,\delta) = (\beta-\delta)^2$, is given by,

$$
\begin{aligned}
\hat{\beta} &= E(\beta|\underline{x}) \\
&= \int_0^\infty \beta f(\beta|\underline{x}) d\beta \\
&= \int_0^\infty \beta \frac{(S+a)^{n+b} e^{\frac{-(S+a)}{\beta}}}{\beta^{n+b+1}\Gamma(n+b)} d\beta \\
&= \frac{n\bar{X}+a}{n+b-1}.
\end{aligned}
$$

3.5 Exponential Distribution(two-parameter)

In this section we will discuss the estimation problem for two-parameter exponential distribution.

Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed random samples taken from two-parameter exponential distribution $Ex(\alpha, \beta)$. We know the method of moments of estimator as well as the maximum likelihood estimator of the parameter (α, β) are

$$
\hat{\alpha} = \bar{X} - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,
$$

$$
\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.
$$

Now we try to get the Bayes estimator of the parameters (α, β).
The probability density function of the random variable X is given by,

$$
f(x;\alpha,\beta) = \frac{1}{\beta}e^{-(x-\alpha)/\beta};\ \alpha < x < \infty,\ -\infty < \alpha < \infty,\ \beta > 0.
$$

The likelihood function is given by,

$$
\begin{aligned}
L(\underline{X};\alpha,\beta) &= \prod_{i=1}^{n}\frac{1}{\beta}e^{-(x-\alpha)/\beta} \\
&= \frac{1}{\beta^n}e^{-\frac{1}{\beta}\sum_{i=1}^{n}(X_i-\alpha)} \\
&= \frac{1}{\beta^n}e^{\frac{-1}{\beta}\{S+n(X_{(1)}-\alpha)\}}.
\end{aligned}
$$

where $X_{(1)}$ is the first order statistic in the sample. Here $\underline{X} = (X_1, X_2, \ldots, X_n)$ and $S = \sum_{i=1}^{n}(X_i - X_{(1)})$.

The joint pdf is given by,

$$\begin{aligned}\Pi(\underline{X}; \alpha, \beta) &= L(\underline{X}; \alpha, \beta) g(\alpha, \beta) \\ &= \frac{1}{\beta^{n+1}} e^{\frac{-1}{\beta}\{S+n(X_{(1)}-\alpha)\}}.\end{aligned}$$

The marginal pdf of α is given by,

$$\begin{aligned}\Pi_1(\alpha|\underline{X}) &= \int_0^\infty \Pi(\underline{X}; \alpha, \beta) d\beta \\ &= \frac{k}{[S + n(X_{(1)} - \alpha)]^n} \;, \quad -\infty < \alpha < X_{(1)},\end{aligned}$$

where

$$k^{-1} = \int_{-\infty}^{X_{(1)}} \frac{d\alpha}{[S + n(X_{(1)} - \alpha)]^n}.$$

Let

$$S + n(X_{(1)} - \alpha) = V,$$

then

$$d\alpha = -dV/n.$$

Finally we get,

$$k^{-1} = \frac{1}{n(n-1)S^{n-1}}.$$

Substituting k in $\Pi_1(\alpha|\underline{x})$ we have

$$\Pi_1(\alpha|\underline{X}) = \frac{n(n-1)S^{n-1}}{[S + n(X_{(1)} - \alpha)]^n}.$$

Now the Bayes estimator of α with respect to the squared error loss function is given by,

$$\begin{aligned}\hat{\alpha} &= E(\alpha|\underline{X}) \\ &= \int_{-\infty}^{X_{(1)}} \alpha \frac{n(n-1)S^{n-1}}{[S + n(X_{(1)} - \alpha)]^n} d\alpha.\end{aligned}$$

After certain calculations we get,

$$\hat{\alpha} = X_{(1)} - \frac{S}{n(n-2)}.$$

The marginal pdf of β is given by,

$$\begin{aligned}\Pi_2(\beta|\underline{X}) &= \int_{-\infty}^{X_{(1)}} \Pi(\underline{X};\alpha,\beta)d\alpha \\ &= \frac{S^{n-1}}{\Gamma(n-1)\beta^n}e^{-S/\beta}.\end{aligned}$$

The Bayes estimator of β with respect to the squared error loss function is given by,

$$\begin{aligned}\hat{\beta} &= E(\beta|\underline{X}) \\ &= \frac{S}{n-2}\ ,\ n > 2.\end{aligned}$$

Chapter 4

Bayesian Estimation Using Lindley's Approximation

4.1 Introduction

In this chapter, we will take up a problem of estimating parameter of some well known distribution functions, such as binomial, Poisson, normal and exponential distribution functions using Lindley's approximation.

The basic idea in Lindley's approach is to obtain Talyor series expansion of the function involved in posterior moment,

$$E\left\{u(\theta)|\underline{x}\right\} = \frac{\int_\Omega u(\theta)v(\theta)e^{L(\theta)}d\theta}{\int_\Omega v(\theta)e^{L(\theta)}d\theta}.$$

where $u(\theta)$ and $v(\theta)$ are arbitrary functions of θ and Ω is the parameter space of θ. $L(\theta)$ is the logarithm of likelihood function. About the maximum likelihood estimator $\hat{\theta}$. Lindley approximated that posterior moment by,

$$E\left\{u(\theta)|\underline{x}\right\} = \left[u + \frac{1}{2}\sum_i\sum_j(u_{ij} + 2u_i\rho_j)\sigma_{ij} + \frac{1}{2}\sum_i\sum_j\sum_k\sum_r L_{ijk}\sigma_{ij}\sigma_{kr}\right] + O\left(\frac{1}{n^2}\right) \quad (4.1)$$

where

$$i, j, k, r = 1, 2, \ldots, m;$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_m),$$

$$\hat{\theta} \text{ is the MLE of } \theta,$$

$$u_i = \frac{\partial u}{\partial \theta_i}, u_{ij} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j}, L_{ijk} = \frac{\partial^3 L}{\partial \theta_i \partial \theta_j \partial \theta_k}$$

$$\rho = \rho(\theta) = \log v(\theta), \rho_i = \frac{\partial \rho}{\partial \theta_i},$$

and σ_{ij} is the $(i,j)^{th}$ element in $[-L_{ij}]^{-1}$.

For a single parameter case $m = 1$, (4.1) reduces to

$$E(u|x) = \left[u + \frac{1}{2}(u_2 + 2u_1\rho_1)\sigma^2 + \frac{\sigma^4}{2}L_3 u_1 + O\left(\frac{1}{n^2}\right)\right]_{\hat{\theta}} \tag{4.2}$$

In the following section 4.2 we will discuss about the binomial parameter, in section 4.3 we will discuss about the Poisson parameter, in section 4.4 we will discuss about the exponential parameter and in section 4.5 we will discuss about the normal parameter for finding the Bayesian estimator using Lindley's approximation.

4.2 Bayesian estimation of binomial parameter using Lindley's approximation

In this, we know the probability mass function is

$$f(x|(n,p)) = \binom{n}{x} p^x q^{n-x}, \; q = 1 - p; \; x = 0, 1, 2, ..., n.$$

We will obtain Bayes estimator of p under AlI prior $v(p) \propto \frac{1}{p(1-p)}$ using the approximation for m=1.

We have $u = p, u_1 = 1, u_2 = 0, \; \rho \propto -[\log p + \log(1-p)]$.

Logarithmic likelihood function for binomial distribution is given by,

$$L = Const + x \log p + (n - x) \log(1 - p).$$

Now calculating

$$\begin{aligned} \frac{\partial^2 L}{\partial p^2} &= -\frac{x}{p^2} - \frac{n-x}{(1-p)^2} \\ L^3 &= \frac{\partial^3 L}{\partial P^3} \\ &= \frac{2x}{p^3} - \frac{2(n-x)}{(1-p)^3} \end{aligned}$$

and

$$\rho_1 = \frac{2p-1}{p(1-p)}.$$

At $p = \hat{p} = \frac{x}{n}$, we have

$$\begin{aligned} \frac{\partial^2 L}{\partial p^2} &= -\frac{n}{p(1-p)}, \\ \sigma^2 &= \frac{p(1-p)}{n} \\ L_3 &= 2n \left[\frac{1}{p^2} - \frac{1}{(1-p)^2} \right] = \frac{2n(1-2p)}{p^2(1-p)^2}. \end{aligned}$$

Substituting in (4.2)

$$\begin{aligned} p^* &= \left[p + \frac{2p-1}{p(1-p)} . \frac{p(1-p)}{n} + \frac{n(1-2p)}{p^2(1-p)^2} . \frac{p^2(1-p)^2}{n^2} + O\left(\frac{1}{n^2}\right) \right]_{p=\hat{p}} \\ &= p + \frac{2p-1}{n} - \frac{2p-1}{n} + O\left(\frac{1}{n^2}\right) \\ &= p + O\left(\frac{1}{n^2}\right) \\ &= \frac{x}{n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

4.3 Bayesian estimator of Poisson parameter using Lindley's approximation

Consider the probability mass function of Poisson distribution is

$$f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Now we will obtained the Bayes estimator of λ under Jeffrey's invariant prior $v(\lambda) = \frac{1}{\lambda}$ using the approximation for m=1. Logarithmic likelihood function for Poisson distribution is given by,

$$L = -n\lambda + \sum_{i=1} nx_i \log\lambda - \log(x_1!x_2!...x_n!).$$

Now calculating

$$\frac{\partial L}{\partial \lambda} = 0 \Longrightarrow -n + \frac{\bar{x}}{\lambda} = 0 \Longrightarrow \lambda = \frac{\bar{x}}{n}$$

$$\begin{aligned} \frac{\partial^2 L}{\partial \lambda^2} &= -\frac{\bar{x}}{\lambda^2} \\ L^3 &= \frac{\partial^3 L}{\partial \lambda^3} = \frac{2\bar{x}}{\lambda^3} \\ \rho &= \log\lambda \\ \rho_1 &= \frac{\partial \rho}{\partial \lambda} = -\frac{1}{\lambda}. \end{aligned}$$

Taking $u = \lambda,\ u_1 = 1,\ u_2 = 0$

$$\sigma^2 = \left[-\frac{\partial^2 L}{\partial \lambda^2}^{-1} \right] = \frac{\lambda^2}{\bar{x}} = \frac{\bar{x}}{n^2}.$$

The Bayes estimator of λ is given by substituting the above in (4.2) we get

$$\begin{aligned} \lambda^* &= \frac{\bar{x}}{n} + \frac{\bar{x}}{\lambda n^2} + \frac{\bar{x}^3}{\lambda^3 n^4} + O\left(\frac{1}{n^2}\right) \\ &= \frac{\bar{x}}{n} - \frac{1}{n} + \frac{1}{n} + O\left(\frac{1}{n^2}\right) \\ &= \frac{\bar{x}}{n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

4.4 Bayesian estimation of exponential parameter using Lindley's approximation

Consider the exponential probability density function

$$f(x|\theta) = \frac{1}{\theta} e^{\left(-\frac{x}{\theta}\right)}, x,\ \theta > 0.$$

We derive Bayes estimator of θ with respect to ALI prior $v(\theta) \propto \frac{1}{\theta^2}$ using the approximation.

We have $u = \theta,\ u_1 = 1,\ u_2 = 0,\ \rho \propto -2\log\theta,$

$$\begin{aligned} \rho_1 &= \frac{-2}{\theta} \\ L &= -n\log\theta - \frac{n\bar{x}}{\theta} \\ L_2 &= \frac{n}{\theta^2} - \frac{2n\bar{x}}{\theta^3} \\ L_3 &= \frac{2n}{\theta^3} + \frac{6n\bar{x}}{\theta^4}. \end{aligned}$$

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

$$L_2 = \frac{-n}{\theta^2},\ \sigma^2 = \frac{\theta^2}{n}$$

and

$$L_3 = \frac{4n}{\theta^3}.$$

Substituting in (4.2) we get

$$\begin{aligned} \theta^* &= \left[\theta - \frac{2\theta^2}{n\theta} + \frac{2n\theta^4}{n^2\theta^2} + 0\left(\frac{1}{n^2}\right)\right]_{\hat{\theta}} \\ &= \hat{\theta} + 0\left(\frac{1}{n^2}\right) \\ &= \bar{x} + 0\left(\frac{1}{n^2}\right). \end{aligned}$$

4.5 Bayesian estimation of normal parameter using Lindley's approximation

Considering the probability density function of normal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x, \mu < \infty, \ \sigma > 0.$$

Then the likelihood function is

$$l(\mu,\sigma|\bar{x}) = \frac{k}{\sigma^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

$$\begin{aligned} L = \log l &= k \log \frac{1}{\sigma^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \\ &= \log k + \log \frac{1}{\sigma^n} + \log e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \\ &= constant - n \log \sigma - \frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

Here we consider the two parameter case, $m = 2$, $\theta = (\theta_1, \theta_2)$ From (4.1) we have

$$\begin{aligned} u^* &= [u + \frac{1}{2}(u_{11}\sigma_{11} + u_{22}\sigma_{22}) + \rho_1 u_1 \sigma_{11} + \rho_2 u_2 \sigma_{22} \\ &+ \frac{1}{2}\{\sigma_{11}\sigma_{22}(u_1 L_{12} + u_2 L_{21}) + u_1 \sigma_{11}^2 L_{30} + u_2 \sigma_{22}^2 L_{03}\} + O(\frac{1}{n^2})]_{\hat{\theta}} \end{aligned}$$

Now calculating

$$\begin{aligned} L_{30} &= \frac{\partial^3 L}{\partial \mu^3} = 0, \\ L_{12} &= \frac{\partial^3 L}{\partial \mu \partial \sigma^2} = \frac{6\sum_i (x_i - \mu)}{\sigma^4}, \\ L_{21} &= \frac{\partial^3 L}{\partial \mu^2 \partial \sigma} = \frac{2n}{\sigma^3}, \\ L_{03} &= \frac{\partial^3 L}{\partial \sigma^3} = \frac{-2n}{\sigma^3} + \frac{12}{\sigma^5}\sum_i (x_i - \mu)^2, \\ \frac{\partial^2 L}{\partial \mu^2} &= \frac{-n}{\sigma^2}, \ \frac{\partial^2 L}{\partial \mu \partial \sigma} = \frac{-2}{\sigma^3}\sum_i (x_i - \mu), \\ \frac{\partial^2 L}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_i (x_i - \mu)^2. \end{aligned}$$

Since μ and σ are orthogonal to each other $\theta_{ij} = 0, \; i = j, \; \theta_1 = \mu, \; \theta_2 = \sigma$.
At the MLE $(\hat{\mu}, \hat{\sigma})$ we have

$$[-L_{ij}] = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

then

$$\sigma_{11} = \frac{\sigma^2}{n}, \qquad \sigma_{22} = \frac{\sigma^2}{2n},$$

and

$$\begin{aligned} L_{30} &= 0, \\ L_{03} &= \frac{10n}{\sigma^3}, \\ L_{21} &= \frac{2n}{\sigma^3}, \\ L_{12} &= 0. \end{aligned}$$

Substituting in (4.1) we get

$$\begin{aligned} \mu^* &= \bar{x} + O\left(\frac{1}{n^2}\right) \\ \sigma^* &= \left[\sigma + \frac{\sigma^2}{2n}\left(\frac{-1}{\sigma}\right) + \frac{1}{2}\left(\frac{\sigma^4}{2n^2} \cdot \frac{2n}{\sigma^3} + \frac{\sigma^4}{4n^2} \cdot \frac{10n}{\sigma^3}\right) + O\left(\frac{1}{n^2}\right)\right]_{\hat{\sigma}} \\ &= \hat{\sigma}\left(1 + \frac{5}{4n}\right) + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Similarly, if we take the parameter $u(\theta) = \mu + \eta\sigma$
then $u_1 = 1, \quad u_2 = \eta, \quad u_{11} = 0, \quad u_{22} = 0$
Hence

$$\begin{aligned} \mu^* &= \bar{x} + O\left(\frac{1}{n^2}\right) \\ \sigma^* &= \sigma - \frac{\eta\sigma}{2n} + \frac{\eta\sigma}{2n} + \frac{5\eta\sigma}{4n} + O\left(\frac{1}{n^2}\right) \\ &= \hat{\sigma}\left(1 + \frac{5\eta}{4n}\right) + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Conclusions and Scope of Future Works

In this project work, at first the basic requirements for the development of subsequent chapters are studied. We have learned some techniques for estimating parameters of a distribution function such as maximum likelihood, method of moments and the Bayesian approach. In particular the problem has been considered for the case of binomial, Poisson, normal and exponential distribution. We have also studied some characteristics of the estimators. Some of the future works to be carried out are listed below.

- Taking various types of prior we can get Bayesian estimator of different types of distribution functions.
- Using Lindley's approximation we can get Bayesian estimator of some complicated integrals whose closed form are not possible.
- Bayes estimator with respect to different loss functions can be obtained.

Bibliography

[1] Berger, James, O. (1985): Statistical Decision Theory, Foundations, Concepts and Methods. Springer-Verlag, New York.

[2] Johnson, N. L. , Kotz, S. and Balakrishnan, N.(2004): Continuous Univariate Distributions Vol-1, Wiley-interscience, New York.

[3] Johnson, N. L. , Kotz, S. and Balakrishnan, N.(2004): Continuous Univariate Distributions Vol-2, Wiley-interscience, New York.

[4] Lehman, E. L. and Casella, G. (2008): Theory of point Estimation, Springer, New York.

[5] Lindley, D. V. (1980): "Approximate Bayesian Method", Trabajos Estadistica, 31, 223-237.

[6] Miller, M. (2003): John E.Freund's Mathematical Statistics, Prentice-Hall of India, New Delhi.

[7] Rohatgi, V. K. and Saleh, A. K. Md. E. (2009): An Introduction to Probability and Statistics, Wiley-interscience, New York.

[8] Sinha, S. K. (2004): Bayesian Estimation, New Age International, New Delhi.