Chapter 1

Introduction: Definition and Terminologies

1.1 Introduction

We are frequently reminded of the fact that we are living in the information age. The information age is contemporary to the modern age of technological development and ecosystem. In the modern age of technological development, we have the requirement of ecological resources and technical items/devices to access the knowledge, benefiting in our daily life uses. The lifetime and ecological literature have adequacy to continue further study. Therefore, these fields are very attractive to draw the attention of researchers. In this context, we are focusing on lifetime and ecological study.

We make different choices in our lives to make our lives more convenient, to perform our duties more efficiently, and for a variety of purposes. However, statistics, with its many useful methods, aids us in making decisions under uncertainty, and as a result, it is gradually being embraced as the science of decision-making by many branches of science, applied disciplines, social sciences, and even literature. In statistics, uncertainty is defined as randomness, which is calculated in terms of probability. This theory serves as the foundation for a study in which

decisions are taken based on a small number of data points (referred to as a sample) drawn from the population of interest. Variables such as human survival times and lifetime of other man-made systems/items, customer waiting times in line to receive service at a center, and other economic and demographic variables are often encountered in real-life situations. The term "lifetime" refers to the length of time it takes for an object to enter its failed state. It is worth noticing that all of these variables have the same level of support, namely the positive real-line number. Since statistical studies of such lifetimes play such a special and significant role in statistical procedures, they are classified as a distinct branch of statistics known as Survival/Reliability analysis. As a result, survival analysis is a technique for analyzing failure (death) time results (time to event data). Medicine, genetics, public health, epidemiology, engineering, economics, and demography are a few of the areas where it can be used.

Also, our daily life activity goes with the environment and ecology. The term ecology was coined by Ernst Haeckel in 1869. Ecology is the branch of science in which we study the presence of species in an organism and its environment, including individual habitat, population, community, ecosystem, and biospheres as a whole. Ecology is the study of households with emphasis on the totality or pattern of relationship between living organisms to one another and to their surroundings, their natural environment, and ecosystems. Ecology is defined as the study of an ecosystem. These household consists of non-living matter such as soil, water, light, wind, humidity, minerals, gases, etc., and living organisms such as micro-organisms, plants, animals, bacteria, and humans. An organism depends upon each other for its survival, existence, and continuance. Besides, living (biotic) organisms and their non-living (abiotic) environment are inseparably interrelated and interact with each other, see Dash (2001). It is a fascinating discipline because everyone is usually interested in knowing about his surroundings. Ecology is concerned with the biology of organisms, population, communities, etc., and their functional processes occurring in natural habitats like ponds, lakes, oceans, and land. A community or biotic community includes all the population of a given area, called the habitat. The community and the abiotic environment interact and function together as a system called the "ecological system" or "ecosystem", a term coined by the British ecologist Arthur Tansley in 1935. Tansley

devised the concept to draw attention to the importance of transfers of materials between organisms and their environment. He regarded ecosystems not simply as natural units, but as mental isolates and later defined the spatial extent of ecosystem using the term ecotope. The functional form of the ecosystems is of great concern to ecologists. Ecosystems show large variations in their size, structure, composition, and so on. However, all the ecosystems are characterized by certain basic structural and functional features which are common. There can be different types of ecosystems such as forest ecosystem, desert ecosystem, and marine ecosystem.

Species richness estimation can be extended to both animals and organisms in biology. The term "species richness" is used to describe the number of species that live in a given biosphere or population. The number of species in an ecosystem will help assess its complexity. Finding a high level of species richness can aid in the identification of populations that have been under-sampled. The number of endangered or extinct organisms can be calculated using measurements taken over time. Species abundance is the number of individuals per species and relative abundance refers to the evenness of distribution of individuals among species in a community. Two communities may be equally rich in species but differ in relative abundance. Animal populations of interest can range from very large animals, such as whales Zeh et al. (1986), Raftery and Zeh (1998), to bacteria that can only be observed under a microscope Hong et al. (2006). Other interesting animal populations for which diversity is studied are fish Smith and Jones (2005), fossils Cobabe and Allmon (1994), and birds Borgella Jr and Gavin (2005); Walther and Martin (2001).

Initially, two papers based on estimating the number of classes was written by Good (1953) and Fisher et al. (1943), with the interest of estimating the frequencies of species in an animal population. Good (1953) estimates the probability of an unseen species as $n_1 = n$ where n_1 is the number of species represented by only one individual in the sample and n is the total number of observed individuals. Fisher et al. (1943) model the species abundances with a parametric gamma-mixed Poisson or negative binomial distribution. The negative binomial model is based on assuming that the numbers of individuals from each species are independent Poisson samples and that the means of these Poisson random variables follow a gamma distribution. Many other

approaches, including Bayesian methods, have been developed for the species problem since these early works. For a review on this problem including other related models and additional applications see Bunge and Fitzpatrick (1993), Buckland et al. (2000), Pollock (2000), Schwarz and Seber (1999), and Seber and Schwarz (2002).

1.2 Classical Inference

In the classical inference, a population have some characteristic of the elements that can be represented by a RV X whose density is $f(X,\theta)$, where the form of the density is assumed to be known except that it contains an unknown parameter θ and make inferences about θ based on information contained in the observed sample only.

The classical school believes in *Fisher's 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 *Fisher's* ML estimator. In spite of certain limitations, the ML estimators have a number of desirable properties and are extensively used in preference of the other classical estimators.

In ML estimation method, it seems that a good estimate of the unknown parameter θ would be the value of θ that maximizes the likelihood function. Parameter θ may be discrete and continuous. Suppose we have a random sample $(x_1,x_2,x_3,\ldots,x_n)$ for which the PDF of x_i is $f(x_i,\theta)$. Then the joint PDF or PMF of $(x_1,x_2,x_3,\ldots,x_n)$ is denoted by $L(x,\theta)$ as,

$$L(x,\theta)=\prod_{i=1}^{n} f(x_i,\theta). \tag{1.1}$$

Likelihood function in Equation (1.1) can be maximized through LS method for continuous parameter and LDS method for discrete parameter, see (Jain et al. (2003) and Lindsay and Roeder (1987)). There are many different methods of point estimation, mentioned as method of moment, method of least square (LS), and methods based on quantile/percentile, etc., are discussed in the literature. But the ML estimator is very popular and widely used since it

has many optimum properties as consistency, invariance, and other asymptotic properties, see Rohatgi and Saleh (2015).

Point estimation, one may be interested in finding a set of values, say $A(\theta)$ such that $A(\theta)$ contains the true value of the parameter (θ) with a certain high probability $(1-\phi)$, $\phi \in (0,1)$. Let $T_1 = t_1(X_1, X_2, X_3, \cdots, X_n)$ and $T_2 = t_2(X_1, X_2, X_3, \cdots, X_n)$, $T_1 \leq T_2$ be two statistics such that

$$P_\theta[T_1 \leq \theta \leq T_2] = 1-\phi, \forall \theta \in \Theta,$$

where $(1-\phi)$ does not depend on θ. Then the random interval (T_1, T_2) is called the $100(1-\phi)\%$ confidence interval (CI) for θ. Exact CIs are generally not available for some distribution especially in the case of the multi-parameters model. In such a case, that the asymptotic distribution property of ML estimator is very useful in constructing asymptotic CIs principle. The asymptotic CIs are defined as

$$\{\hat{\theta} \mp Z_{\phi/2}\sqrt{var(\hat{\theta})}\}, \tag{1.2}$$

where, $\phi/2^{th}$ upper percentile of a standard normal variables, and $var(\hat{\theta})$ is asymptotic variance of θ, and is obtained as diagonal elements of inverse fisher's information matrix. Suppose there are k parameters $\theta = \theta_1, \theta_2, \cdots, \theta_k$, then variance-covariance matrix, say V, is defined by

$$V(\theta) = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots \delta_{1k} \\ \delta_{21} & \delta_{22} & \cdots \delta_{2k} \\ . & . & . \\ . & . & . \\ . & . & . \\ \delta_{k1} & \delta_{k2} & \cdots \delta_{kk} \end{bmatrix}^{-1}_{\theta=\hat{\theta}}$$

where, $\delta_{ij} = -\frac{\partial^2}{\partial\theta_i, \partial\theta_j} logL; (i,j) = 1,2,3,\cdots,k$, second derivatives of the likelihood function with respect to θ.

1.3 Bayesian Inference

In Bayesian inference, it is believed that we always have the availability to make use of subjective probabilities that measures the degrees of belief about the value or values of unknown parameter θ. These subjective probabilities are used to define prior distribution for the parameter θ, prior to sampling. In other words, the parameter θ may be treated as a RV with known prior distribution, say $g(\theta)$, see Berger (2013). The ML method, as well as other classical approaches, are based only on the empirical information provided by the available data. However, when there are some technical knowledge on the parameters of the distribution available, a Bayesian inference seems to be an attractive inferential method.

The goal of Bayesian inference is to represent prior uncertainty about model parameters with a probability distribution and to update this prior uncertainty with current data to produce a posterior probability distribution for the parameter that contains less uncertainty. The posterior distribution, denoted by $\pi(\theta|x)$ of the parameter, say θ given x is defined to be the conditional distribution of θ given the sample observations x and is given by

$$\pi(\theta|x) = \frac{g(\theta)L(x|\theta)}{\int_{\Theta} g(\theta)L(x|\theta)d\theta}, \tag{1.3}$$

where $g(\theta)$ is the prior distribution function that reflects beliefs about θ (prior to experimentation) and Equation (1.3) shows the updated belief about θ after observing the sample.

Clearly, the prior distribution plays a very crucial role in Bayesian of parameter. But the specification of the prior distribution is not an easy task. A vast Bayesian literature is fully devoted to finding the prior distribution of the parameter of interest, see for more details on prior specifications Jeffreys (1946), Zellner (1986a), Box and Tiao (2011), Berger (2013), Gelman et al. (2013). In the next section, we will discuss prior distributions.

1.4 Prior Distribution

The prior distributions are broadly classified into two categories, (i) Informative and (ii) Non-informative. Berger (2013) has given a useful discussion on various methods to construct a prior distribution is certainly a subjective manner. The main idea of subjective probability is to let the probability of an event reflect the personal belief in the chance of occurrence of the event, however, it is typically determined by introspection. The simplest way of determining subjective probabilities is to compare events, determining relative likelihoods. Several useful techniques are included (i) The Histogram Approach, (ii) The Relative Likelihood Approach, (iii) CDF Determination Approach, and (iv) Matching a Given Functional Form. The easiest and very simplified approach is to use a given functional form when a piece of prior information is available, and the problem is then reduced to a subjective determination of a few hyper-parameters (see Good (1950)).

For example, a gamma distribution is widely discussed in existing literature as a prior distribution for the parameters of numerous lifetime models. The gamma distribution as a prior have shape parameter α and scale parameter β for the parameter θ is defined as

$$g(\theta) = \frac{\alpha^{\beta}}{\gamma(\beta)}\theta^{\beta-1}e^{-\alpha\theta}; \qquad \theta > 0, \alpha > 0, \beta > 0. \tag{1.4}$$

Here, in the Bayesian paradigm, the parameters (α, β) of the prior distribution are referred to as hyper-parameters and are assumed to be known. The simplest way of eliciting these hyper-parameters is to set the prior moments equal to the values taken by the experimenter to guess the prior parameters. In the above case, the moments equations are

$$\frac{\beta}{\alpha} = M(say), \tag{1.5}$$

$$\frac{\beta}{\alpha^2} = V(say), \tag{1.6}$$

where the quantities M and V reflect the experimenter's beliefs about the mean and variance of the unknown parameter of the model. Note that the smaller, moderate and larger values of V respectively correspond to the low, moderate, and high magnitude of beliefs in the expected value of the parameter. The large values of V lead to a flat prior density whereas smaller values of V indicate a high picked prior density for θ. This facile methodology has been widely adopted by many authors for prior elicitation under Bayesian paradigm, see Kundu and Howlader (2010) and Singh et al. (2013a).

A large part of Bayesian literature is devoted to finding appropriate prior distributions for which $\pi(\theta|x)$ can be easily calculated. These are so-called conjugate priors and were developed extensively by Schlaifer and Raiffa (1961). Also defined is a family of prior distributions known as conjugate prior which eases the associated computational difficulties with Bayesian analysis. A prior is said to be conjugate if it's resulting posterior also belongs to the same family of distributions as that of the prior. A number of informative priors such as g-prior by Zellner (1986b), an independent t-gamma prior by Leamer and Leamer (1978) and maximum entropy prior by Berger (2013) have also been suggested in Bayesian literature.

A general class of prior that is often considered in Bayesian analysis is popularly known as non-informative priors. The non-informative priors are those priors that utilize very minimal or no prior information in their choice. Uniform and Jeffery's priors (Jeffreys (1946)) are widely used non-informative priors. If a non-informative prior density is desired, it seems reasonable to give equal weight to all possible values of θ, arriving at the uniform non-informative prior $g(\theta) \equiv c$. Although this was routinely done by Laplace (1812), it came under severe (though unjustified) criticism because of a lack of invariance property under transformation, see Jaynes (1983). The lack of invariance of the constant prior has led to a search of non-informative priors which are appropriately invariant under transformation. Efforts to derive non-informative priors through consideration of transformation of a problem had their beginnings with Jeffreys (1961). It has extensively used in Hartigan et al. (1964), Jaynes (1968), Jaynes (1983), Villegas (1977), Villegas (1981) and elsewhere. A very useful prior is Jeffrey's prior (Jeffreys (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 $I(\theta)$, which is

$$g(\theta) = |I(\theta)|^{1/2}, \tag{1.7}$$

where $|.|$ denotes the determinant and Fisher information matrix under commonly satisfying assumptions (see Lehmann (1983)) this is given by

$$I(\theta) = -E_{\theta}\left[\frac{\partial^2 \log L(x|\theta)}{\partial \theta^2}\right]. \tag{1.8}$$

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 $f(x|\theta)$.

Another non-subjective prior is reference prior, introduced by Bernardo (1979) and further developed by Berger and Bernardo (1989), Berger and Bernardo (1992a), Berger and Bernardo (1992b) is, to the best of our knowledge, the only available method to derive non-subjective posterior distributions which satisfy all these property, (i) Invariance (ii) Consistent marginalization (iii) Consistent sampling properties (iv) Generality (v) Admissibility. Reference prior is non-informative prior and it also has improper nature. Reference prior is an objective prior which is based on maximizing the expected entropy provided by the prior. The amount of information to be expected from an experiment about some quantity of interest naturally depends on the available prior knowledge: the more prior information available, the less information may be expected to be learned from the data. An infinitely large experiment would eventually provide all missing information; thus, it is possible to obtain a measure of the amount of missing information as a limiting form of a functional of the prior distribution.

1.5 Empirical Bayesian Inference

A major drawback of the conventional approach is that it cannot use past data, while the Bayesian approach is that it requires an assumption for the prior distribution. It appears to be desirable to have an approach, which can make use of the past data and does not require an assumption for the prior distribution. It means we did not know the value of prior parameters. Many statisticians have considered working in areas where it is possible to specify prior distributions reasonably well. Among the others, empirical Bayesian have typically assumed that the parameters of interest come from some unknown prior distributions instead of specifying the prior distributions in advance. They have developed the technique, called the Empirical Bayesian approach, for eliciting the unknown prior distributions using the past available sample information. Borrowing the strengths of each, Herbert Robbins proposed an approach, known as empirical Bayesian approach, as a valid alternative to make use of past data to estimate the statistical form of prior information. Empirical Bayesian methods comparatively require fewer sample data to achieve the same quality of inference than the methods based on sampling theory. This is one important consideration when sample data is either expensive or difficult to obtain. Thus, the Empirical Bayesian approach sounds like a compromise between the conventional approach and the Bayesian approach. It is described extensively in the literature, e.g., Robbins (1955), Robbins (1964) and Maritz (1967), Sinha et al. (1976), Grabski and Sarhan (1996), Ahmad et al. (1997), Pensky and Singh (1999), Jaheen (2004). Recently, some more developments are Shojaee et al. (2012) that discuss the Empirical Bayes estimators for compound Rayleigh distribution parameter and reliability under record data.

Suppose that θ is a vector, consisting of components $(\theta_1, ..., \theta_p)$ that are i.i.d. from the density π_0. Suppose also that the data X consists of independent components $(X_1, X_2, ..., X_p)$ where each X_i has density $f(x_i|\theta_i)$. Then the common marginal distribution of each X_i is

$$m_0(x_i) = \int f(x_i|\theta_i) dF^{\pi_0}\theta_i, \tag{1.9}$$

and $(X_1, X_2, ..., X_p)$ can be considered to be a simple random sample from m_0. Note that this also follows from the direct calculation (assuming continuous densities for convenience)

$$\begin{aligned} m(x) &= \int f(x|\theta)\pi_0(\theta)d\theta \\ &= \int \left[\prod_{i=1}^{p} f(x_i|\theta_i)\right]\left[\prod_{i=1}^{p}\pi_0(\theta_i)\right]d\theta_i \\ &= \prod_{i=1}^{p}\int f(x_i|\theta_i)\pi_0(\theta_i)d\theta_i \\ &= \prod_{i=1}^{p} m_0(x_i). \end{aligned}$$

The data x can thus used to estimate m_0 (and hence m).

1.6 Hierarchical Bayesian Inference

An another important type of prior distribution is a hierarchical prior, also called a multistage prior, see Lindley and Smith (1972). The idea is that one may have structural and subjective prior information at the same time, it is often convenient to model at this stage. The hierarchical approach is most commonly used when the first stage, Γ, consists of prior of a certain functional form see Antoniak (1974), Berry et al. (1979), and Kuo (1986) being exceptions. Thus, if

$$\Gamma = \{g_1(\theta|\lambda) : g_1 \; is \; of \; a \; given \; functional \; form \; and \; \lambda \in \Lambda\}, \tag{1.10}$$

then the second stage would consist of putting a prior distribution, $g_2(\lambda)$, on the hyper parameter λ, which could be chosen for the hyper parameters according to subjective beliefs. Such a second stage prior is sometimes called a hyper prior for this reason. The difficulty of specifying the second stage prior has made common use of non-informative priors at the second stage. Note that there is no theoretical reason for limiting hierarchical priors to just two stages, but more than two are rarely useful in practice. see also Goel and Degroot (1981) and Goel (1983). As a final comment, note that a hierarchical structure is merely a convenient representation for

a prior, rather than an entirely new entity: any hierarchical prior can be written as a standard prior. For instance, in this situation, the actual prior distribution is

$$g(\theta) = \int_{\Lambda} g_1(\theta|\lambda) dF^{g_2}(\lambda), \tag{1.11}$$

and any Bayesian analysis will actually be performed with respect to g. Attention will be restricted to two-stage priors. The first stage prior, $g_1(\theta|\lambda)$ where λ is a hyper parameter in Λ, can be thought of as the unknown prior in the empirical Bayes scenario. Instead of estimating λ, as in empirical Bayes analysis, however, λ will be given a second stage prior distribution $g_2(\lambda)$. This could be a proper prior but is often chosen to be a suitable non informative prior. It is frequently used to calculate and to write $\lambda = (\lambda^1, \lambda^2)$, represented g_2 as

$$\begin{aligned} g_2(\lambda) &= g_{2.1}(\lambda^1|\lambda^2) g_{2.2}(\lambda^2), \\ g(\theta|x) &= \int_{\Lambda} g_1(\theta|x,\lambda) g_{2.1}(\lambda^1|x,\lambda^2) g_{2.2}(\lambda^2|x) d\lambda. \end{aligned}$$

This process is called Hierarchical Bayesian process for estimating the hyper parameter, see Berger (2013).

1.7 E-Bayesian Inference

In the Bayesian inference, Posterior distribution is the basis of prior distribution and likelihood of the experiment. It depends on the selection of prior distribution and specification of loss functions. But, prior distribution may depend on the prior parameters i.e. hyper parameters. Hierarchical prior distribution has been used as prior for the unknown hyper parameters. Very first time Lindley and Smith (1972) introduced the idea of the hierarchical prior distribution. The hierarchical Bayesian inference have the requirement of prior at least two stages to finish the setting of the prior distribution. Hence, it is more robust as well as more efficient than Bayesian inference. The method for construction of hierarchical prior distribution has been

developed by Han (2007). This method is known as E-Bayesian (Expected-Bayesian) estimation. The hierarchical Bayesian methods have confronted complicated integration, though some computational methods are available in the literature. However, it was observed that obtaining the E-Bayesian estimates of the unknown parameters is simpler than that of hierarchical Bayesian estimates, see Han (2007), Han (2011a), Han (2011b). Further, E-Bayesian estimation for the parameters of different lifetime distributions has been discussed by several authors, see Gupta (2017), Yousefzadeh (2017), Han (2017a), Han (2017b), Han (2019b) etc., and some authors also discussed the E-Bayesian estimation for parameters of lifetime distribution with type-II censoring, see Jaheen and Okasha (2011), Okasha (2014), Reyad and Ahmed (2016) etc. Recently, El-Sagheer (2017) has considered the Rayleigh distribution for E-Bayesian estimation under progressive type-II censoring and Kızılaslan (2017) discussed the hierarchical and E-Bayesian Bayesian estimations for the proportional reversed hazard rate model based on record values. Some recent literatures of E-Bayesian estimation to develop the E-posterior and E-MSE method, see Han (2018), Han (2019a), Han (2019b), Han (2019c) and Han (2020).

1.8 Loss Function

Statistical inference can always be viewed as a decision problem under prevailing uncertainties modeled in the form of parent population distribution. The statistical decisions are based on the sample information only in classical inferences; whereas in Bayesian decisions, in addition, to sample information, it also includes prior information. The overall purpose of statistical inference is to provide an optimal decision based on some evaluation criterion for the goodness of the decisions. The criterion assesses the average consequences of each decision. It is worthwhile to mention here that in statistical inference problems, we try to reveal the true state of affairs therefore two situations may arise. The first one is that the inference will be able to reveal the truth. In this case we are achieving what is intended. Thus there is no loss. In other cases, our inference may deviate from what the truth is and this leads to a loss. In other words, in developing the inferential procedures one should keep this point in mind by specifying an

appropriate loss function. It may be noted that the considered loss functions should be non-negative functions as there is no chance of a gain. A statistical decision problem is formalized by specifying the set of elements (Ω, A, L), where, Ω is the set of all possible values of the parameter called as parameter space, A is set of all action (decision) that taken by statistician and L is the loss function which is a real-valued function of the decision and true state of nature. As mentioned earlier, the statistical decisions are based on the sample information, thus it may be defined as the function from sample space to action space. Thus, our aim is to select from the set of all possible decisions (called decision space), a decision (called optimal decision) for which the average loss (In non Bayesian or classical set up, the expected error is termed as risk defined as expected loss) is minimum. For example, in one-dimensional point estimation problems the true state of nature θ, called parameter, is unknown but one can always specify that it belongs to a set of real numbers $\Theta \in R$ takes $A = \Theta$ meaning that the decision rule will output estimates (guesses) of the true θ. Let us consider that θ is a parameter of some distribution $f(x|\theta)$ and suppose that the parameter θ is estimated by the decision rule $\delta(x)$. Then, the quantity $L(\theta, \delta)$ expresses how much wrong estimates are to be penalized. Then, the Bayes estimators δ^* of the parameter θ is defined as the estimator that minimizes the posterior expected loss i. e.

$$E_{\theta|x}[L(\theta, \delta^*)] = \int_{\Theta} L(\theta, \delta^*)\pi(\theta|x)d\theta. \tag{1.12}$$

In the statistical literature, several loss functions have been discussed namely, square error loss function (SELF), absolute loss function, $0-1$ loss function, quadratic loss function, Linear Exponential (LINEX) loss function, and general entropy loss function (GELF). For more detail about these loss functions, see Winkler (1972), Zellner (1996), Basu and Ebrahimi (1991),Calabria and Pulcini (1994), Box and Tiao (2011), Berger (2013) etc. Among the various loss functions, the most popular one is SELF which is initially proposed by Legendre (1806) and Gauss (1855) to develop the LS theory. Later, it was used in estimation problems when unbiased estimators of parameter θ were evaluated in terms of the risk function $R(\theta, \delta)$ which become nothing but the variance of the estimator. It was observed that SELF is a convex loss

function and another beauty of this loss function is that it equally penalizes the overestimation as well as underestimation of equal magnitude. The mathematical form of the weighted quadratic loss function is given as follows

$$L(\theta,\delta) = \psi(\theta-\delta)^2. \tag{1.13}$$

If ψ is the function of θ then, the corresponding loss function is called a weighted quadratic loss function, and if $\psi = 1$ then we have SELF. i.e.

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

Therefore, the Bayes estimate under SELF is

$$E_{\theta|x}[L(\theta,\delta)] = \int_{\Theta} L(\theta,\delta)\pi(\theta|x)d\theta,$$

after simplification, we get

$$\delta = \int_{\Theta} \theta\pi(\theta|x)d\theta, \tag{1.14}$$

here δ is Bayes estimate under SELF i.e posterior mean. SELF have the symmetric, it is justified for o.e. & u.e. with equal seriousness. Similarly, Bayes estimate under absolute loss function is posterior median and under $0-1$ loss function is posterior mode. However, assumption for real situation of symmetric loss function may not be appropriate. In these situations, when o.e. is more serious than u.e. or vice-versa. Then we have lot of asymmetric loss functions that are available in statistical literature.

The LINEX loss function suggested by Varian (1975) has been widely used by several authors, see Zellner (1986a), Schabe (1991), Pandey and Rai (1992), Ahmadi et al. (2005) and Doostparast (2009). This loss function rises approximately exponentially on one side of zero and approximately linearly on the other side. The mathematical form of this loss function is

$$L(\theta,\delta) = \nu_1\{e^{\nu_1(\theta-\delta)-\nu_1(\theta\delta)-1}\}; \qquad \nu_1 \neq 0, \tag{1.15}$$

where, ν_1 is the loss parameter. If $\nu_1 > 0$, then o.e. is more more serious than u.e. & vice versa. Under this loss function, the Bayes estimate is given by following equation

$$B_L(\delta) = -\frac{1}{\nu_1} log[e^{-\nu_1\theta}]. \tag{1.16}$$

Despite the exibility and popularity of the LINEX loss function for the location parameter estimation, it appears to be unsuitable for the scale parameter, a similar comment can be found in Basu and Ebrahimi (1991) and Parsian and Sanjari Farsipour (1993). To provide a better asymmetric loss function for scale parameter, Basu and Ebrahimi (1991) defined a modified LINEX loss function. Calabria and Pulcini (1994) introduced an alternate of the modified LINEX loss function having the following form,

$$L_G(\theta,\delta) \propto \left\{ \left(\frac{\delta}{\theta}\right)^{\nu} - \nu \log\left(\frac{\delta}{\theta}\right) - 1 \right\}; \qquad \nu \neq 0. \tag{1.17}$$

When $\delta = \theta$, it has minimum. This loss function is generalization of the entropy loss function used by several authors Lindley (1980), Zellner (1986a), Dey et al. (1986), Basu and Ebrahimi (1991), Schabe (1991) and Singh et al. (2016), when shape parameter $\nu = 1$. Here, ν involved in above equation as shape parameter and it reflects the departure from symmetry. When $\nu > 0$ o.e. is considered to be more serious than u.e. of equal magnitude & vice versa. The Bayes estimator of θ under general entropy loss will be

$$B_G(\delta) = \left[E_\theta(\delta)^{-\nu}\right]^{-\frac{1}{\nu}}. \tag{1.18}$$

Provided that, $E_\theta(\delta)^{-\nu}$ exists and is finite. It may be noted here that, when $\nu = -1$ the Bayes estimate under GELF coincides with the Bayes estimate under the SELF.

1.9 Credible Intervals

Another approach to inference is to present a CI for model parameter. The Bayesian analog of a classical CI is called credible interval which is defined as

$$\int_{\theta_L}^{\theta_U} \pi(\theta|x)d\theta = 1-\varphi,$$

where, (θ_L, θ_U) is a $100(1-\varphi)\%$ credible interval for θ. Since the posterior distribution is an actual probability distribution on θ, this interval can be stated with a probabilistic statement. This is in contrast to classical CI which can only be interpreted in terms of CP. The equal tail credible interval for parameter θ, can be obtain by solving the following equation

$$\begin{aligned}\int_0^{\theta_L} \pi(\theta|x)d\theta &= \frac{\varphi}{2},\\ \int_0^{\theta_U} \pi(\theta|x)d\theta &= 1-\frac{\varphi}{2}.\end{aligned}$$

In constructing a credible interval for the parameter θ, it is usually desirable to have those values in the interval which are more probable than those not included in the interval, and such an interval is called the highest posterior density (HPD) interval. The $100(1-\varphi)\%$ HPD credible interval (θ_L^h, θ_U^h) for θ must satisfy the following conditions provided the posterior distribution is unimodal and bell-shaped

$$\int_{\theta_L^h}^{\theta_U^h} \pi(\theta_U^h|x)d\theta = 1-\varphi; \qquad \theta, \in \Theta \tag{1.19}$$

where, $\pi(\theta_L^h|x) = \pi(\theta_U^h|x), \theta \in \Theta$.

1.10 Advanced Statistical Computation Technique

The main obstructions to carrying out statistical inferences are the optimization of the likelihood function to obtain the ML estimator and the integration required to compute the Bayes estimator. The optimization of the likelihood function ultimately results in a solution of k likelihood equations, where k is the number of model parameters to be estimated from the data. The usual iterative methods for solving the likelihood equation

$$L'(\hat{\theta}) = \frac{d}{d\theta} L(\theta) \Big|_{\theta=\hat{\theta}} = 0, \tag{1.20}$$

are based on replacing the left side by the linear terms of its Taylor expansion about an approximate solution $\tilde{\theta}$. If $\hat{\theta}$ denotes a root of above equation, this leads to the approximation

$$0 = L'(\theta) = L'(\tilde{\theta}) + (\hat{\theta} - \tilde{\theta}) L''(\tilde{\theta}),$$

and hence to

$$\hat{\theta} = \tilde{\theta} - \frac{L'(\tilde{\theta})}{L''(\tilde{\theta})}. \tag{1.21}$$

The procedure is then iterated by replacing $\hat{\theta}$ with the value $\tilde{\theta}$ of the right side of the above equation, and so on. This method is referred as the Newton-Raphson iterative process, see Kale (1961) Kale (1962), Barnett (1966). Another method called the fixed-point iterative method alternative to Newton-Raphson (NR) is also used by many practitioners for solving the likelihood equations. In this method, likelihood Equation (1.20) is re-rewritten as

$$\theta = h(\theta),$$

in such a way that any solution of the above equation, which is a fixed point of h, is a solution of Equation (1.20). Reaching to the solution, the iterative process proceeds as

- Set initial solution as $\tilde{\theta}$.

- Update the initial $\tilde{\theta}$ as $\theta_{new} = h(\tilde{\theta})$ and set $\tilde{\theta} = \theta_{new}$.
- Repeat the above step till θ_{new} converges to a value $\hat{\theta}$. Now $\hat{\theta}$ is the solution of Equation (1.21).

For more detail on the fixed-point iterative method and NR method, readers may be referred to Jain et al. (2003). Moreover, these iterative methods may be used to solve complicated functions, sometimes, they do not converge. Then the corresponding point at which the convergence is obtained may not be the desired root. To overcome this difficulty, EM algorithm may be a choice, which is simple to apply and have sure convergence with any initial guess.

1.10.1 Expectation-Maximization Algorithm

Dempster et al. (1977) introduced the term EM algorithm to overcome the above difficulties. They synthesized the earlier formulation of this algorithm in many particular cases and presented a general formulation of this method in finding the ML estimates in a variety of problems. The main references for the EM algorithm are Schafer (1997), Little and Rubin (2019), Tanner (2012), McLachlan and Krishnan (2007) etc. In statistical inference, and EM algorithm is a method for finding ML or maximum posterior estimates of parameters in statistical models, where the model depends on unobserved latent variables. EM algorithm is an iterative method that alternates between performing an expectation (E) step, which computes the expectation of the log-L evaluated using the current estimate for the latent variables, and maximization (M) step, which computes parameters maximizing the expected log-L found on the E step. These parameter estimates are then used to determine the distribution of the variables for the next iteration in the next E step. The EM algorithm is an efficient iterative procedure to compute the ML estimate in the presence of missing or hidden data. In comparison to other optimization techniques, it is very simple and converges reliably. In the case of the unimodal and concave likelihood function, the EM algorithm converges to the global maxima from any starting value, see Wu (1983). Here, we have given a short description of the EM Algorithm. The EM algorithm has two main applications. The first case occurs when the data has missing values

due to limitations or problems with the observation process. The second case occurs when the likelihood function can be simplified by assuming that there are additional but missing parameters. With missing values or parameters in the data which is generated by some distribution under the assumption, we call the data, X, the incomplete data. We assume that the complete data, $Z = (X,Y)$ exists with Y being missing data and that a joint density function also exists as follows,

$$p(Z|\theta) = p(X,Y|\theta) = p(Y|X,\theta)p(X|\theta), \tag{1.22}$$

where, θ is a set of unknown parameters from a distribution including a missing parameter. With the density function, we now define the complete-data likelihood as follows

$$L(\theta|Z) = L(\theta|X,Y) = p(X,y|\theta).$$

The likelihood $L(\theta|X)$ is known as incomplete data likelihood function. Since, we have missing data Y, which have unknown distribution by assumption, we can think of $L(\theta|X,Y)$ as a function of a random variable, Y , with constant values, X and θ.

$$L(\theta|X,Y) = f_{X|\theta}(Y).$$

Using the complete-data log-L function with respect to the missing data Y given the observed data X, the EM algorithm finds its expected value as well as the current parameter estimates at the E step and maximizes the expectation at the M step. Now, repeating E-step and M-step then the algorithm is guaranteed to converge to a local maximum of the likelihood function with each iteration increasing the log-L.

Expectation (E) Step: Firstly, we do the expectation of the complete-data log-L function as

$$Q(\theta|\theta^{(i-1)}) = E[\log\{p(X,Y,\theta|X,\theta^{(i-1)})\}],$$

where $\theta^{(i-1)}$ is a set of current parameters estimates that we use to evaluate the expectation and to increase Q with the new θ for optimization. Here, X and $\theta^{(i-1)}$ are known constants

and θ is a variable to be adjusted. Since Y is a missing RV under an assumed distribution, $f(Y;X|\theta^{(i-1)})$, the expectation in the above equation can be written as

$$E[\log\{p(X,Y,\theta|X,\theta^{(i-1)})\}] = \int_{y\in\Theta} \log\{p(X,y|\theta)\} f(y|X,\theta^{(i-1)})dy,$$

where, Θ is the space of values where y can take values on and $f(y|X;\theta^{(i-1)})$ is the marginal distribution of the missing data Y depending on observed data and current parameters.

Maximization (M) Step: At the M step, we maximize the expectation then the E step, that is to find

$$\theta^{(i)} = arg\max_{\theta}\{\theta^{(i)},\theta^{(i-1)}\}.$$

1.10.2 Bayesian Computation

Bayes estimator required some numerical computation of integration since the integrals encountered in Bayesian analysis or inference of the parameters are often intractable and don't possess the analytical solution. For this reason, Bayesian analysis was often doubting before the invention of versatile computing. The developed computer-intensive sampling methods of estimation have revolutionized the application of Bayesian methods, and such methods now offer a comprehensive approach to complex model parameter estimation. The most common and basic approaches to computing the integrals are approximation methods and sampling methods. One of the simplest approximation methods for evaluating the integrals is called as Gauss-quadrature rule, which is stated as a weighted sum of function values at specified points within the domain of integration, see Golub and Welsch (1969), Hildebrand (1974). Although the numerical integration techniques mainly approximate the integrals by polynomials and were efficiently used in a variety of problems, see Smith et al. (1987), Shaw (1988), Tierney (1994), Smith (1991), but these integral techniques are quite complicated and not easy to use in case of higher-dimension scenarios, see Shaw (1988), Smith (1991), etc.

All of these approaches were methods of approximation, and hence formed a foundation for criticizing Bayesian analysis. Of course, it is true that the Bayesian central limit theorem shows that asymptotically most posterior distributions are normal, see Gelman et al. (2013) and the high dimensional integrals were solved by analytical approximation techniques based on normal approximation, see Heyde and Johnstone (1979), etc. The other approximations developed by Lindley (1980) and Tierney and Kadane (1986) received maximum attention in the Bayesian literature, see Singh et al. (2008a), Singh et al. (2008b), Singh et al. (2009) and references cited therein. These approximation methods were widely used, but these had limitations. If the dimension of the parameter to be estimated is high, these methods become unmanageable as clearly mentioned by Smith (1991). Sampling methods constitute an alternative to approximation methods. The logic of sampling is that we can generate/simulate a sample from the distribution of interest and use discrete formulas, applied to these samples to approximate the integrals of interest. The use of simulation methods for approximating integrals of the form

$$E_f[h(x)] = \int h(x)f(x)dx,$$

can be justified as: the above integration can be approximated by

$$\bar{h_n} = \frac{1}{n}\sum_{i=1}^{n} h(x_i),$$

where, $x's$ are i.i.d. sample of size n from the density $f(x)$, since $\bar{h_n}$ converges almost surely (i.e. for almost every generated sequence) to $E_f[h(x)]$ by the strong law of large numbers. This procedure is referred as Monte Carlo (MC) integration method, here MC refers to the random simulation/process.

Bayesian inference has now become closely linked to sampling-based estimation methods. There are various methods that have been suggested for sample generation. The idea of these procedures started with a concept of rejection sampling that provides a general method for simulation from an arbitrary posterior distribution, but it can be difficult to set up since it requires the construction of a suitable proposal density, see Robert and Casella (2013). In the case

of high dimension model, it is suggested to use more advanced MCMC techniques. Robert and Casella (2013) stated that "a MCMC method for the simulation of a distribution f is any method producing an ergodic Markov chain $(X(t))$ whose stationary distribution is f". The MCMC techniques are relatively straightforward for a range of applications, involving sampling from one or more chains after convergence to a stationary distribution that approximates the posterior, see Gilks et al. (1996).

In fact, the development of MCMC sampling methods, coupled with exponential growth in computing capabilities, has made the use of Bayesian statistics more feasible because of their relative simplicity compared with traditional numerical methods. With the advent of MCMC sampling methods, more complicated and realistic applications can be undertaken, and there is no inherent reliance on asymptotic arguments and assumptions. In the early 1990s, however, the MCMC methods became standard for Bayesian analysis, but novices and other applied scientists who are really attracted to Bayesian methods were very curious about the implementation of MCMC methods and how these really work.

1.10.3 Gibbs Algorithm

Geman and Geman (1984), Geman and Geman (1993) proposed an MCMC algorithm known as Gibbs sampling or Gibbs sampler that is being found to be a very useful MCMC technique for Bayesian analysis under the assumption of high-dimensional models. This procedure permits the simulation from a model possessing high-dimensional parameters to be reduced to the simulation for its much simpler and lower dimensional parameter. Thus, one simulates p random variables sequentially from the p univariate conditionals rather than generating a single p-dimensional vector in a single pass using the full joint distribution. Suppose we wish to simulate a sample from a bivariate posterior, $\pi(\alpha,\beta|x)$ distribution. The two-stage $(p=2)$ Gibbs sampler algorithm can be stated as

- Simulate $\alpha_j \sim \pi_1(\alpha|\beta_{j-1},x)$ with $j=1,2,\cdots,N$ and β_0 is the initial value of β.

- Simulate $\beta_j \sim \pi_2(\beta|\alpha_j, x)$, where $j = 1, 2, \cdots, N$.

where $\pi_1(.)$ and $\pi_2(.)$ are the full conditionals, and N is the sample size to be required for posterior analysis. For large N, α_j and β_j converges to their stationary distributions. Gelfand and Smith (1990) reviewed Gibbs algorithm with another sampling approaches by revealing its potential in a wide variety of conventional statistical problems. Casella and George (1992) studied the properties of Gibbs algorithm with its convergence for many practical problems. Since then, this algorithm has been increasingly employed to perform Bayesian analysis of many real life problems, see Smith and Roberts (1993), Tierney (1994), Brooks (1998), Jackman (2000), Upadhyay and Smith (2001) and Upadhyay and Gupta (2010).

1.10.4 Metropolis-Hastings Algorithm

The Metropolis-Hastings (M-H) algorithm is the most popular MCMC method of sampling. This algorithm was initiated by Metropolis et al. (1953) and generalized by Hastings (1970). Like accept-reject sampling, the M-H algorithm needs to choose a proposal density, say $q(y|x)$ for which sample generation is easy to perform, to simulate a Markov chain from the target density, say f. The choice of proposal density $q(.|x)$ can be almost arbitrarily made in that the only theoretical requirements are that the ratio

$$\frac{f(y)}{q(y|x)},$$

is known up to a constant independent of x and that $q(y|x)$ has enough dispersion to lead to an exploration of the entire support of f. A key advantage of this algorithm over other methods of sampling, like inversion and accept-reject methods, is that it will effectively work with multivariate distributions and do not need an enveloping function as in rejection sampling. The M-H algorithm associated with the target density f and the conditional density q produces a Markov chain $(X^{(t)})$ through the following steps.

- Start with a value $x^{(t)}$ such that $f(x^{(t)}) > 0$.

- Generate Y_t $q(y|x^{(t)})$.
- Calculate the ratio

$$\rho(Y_t, x^{(t)}) = min\left(\frac{f(Y_t)q(x^{(t)}|Y_t)}{f(x^{(t)})q(Y_t|x^{(t)})}, 1\right).$$

- Accept $x^{(t+1)}$ as

$$x^{(t+1)} = \begin{cases} Y_t \ with \ probability \ \rho(Y_t, x^{(t)}) \\ x^{(t)} \ with \ probability \ 1-\rho(Y_t, x^{(t)}) \end{cases}$$

The probability $\rho(.|.)$ is the M-H acceptance probability. This algorithm always accepts values Y_t such that the ratio $\frac{f(Y_t)}{q(Y_t|x^{(t)})}$ is increased, compared to the previous value $\frac{f(x^{(t)})}{q(x^{(t)}|Y_t)}$. When the proposal density is symmetric i.e. $q(Y_t|x^{(t)}) = q(x^{(t)}|Y_t)$, the above M-H algorithm is identical with Metropolis algorithm having Metropolis acceptance function is

$$\rho(Y_t, x^{(t)}) = min\left(\frac{F(Y_t)}{f(x^{(t)})}, 1\right). \tag{1.23}$$

This algorithm is also referred to as an independent M-H algorithm. For more technical details about the properties and implementations of the M-H algorithm, readers may be referred to Robert and Casella (2013). There are cases, associated with the multi-parameters model, in which conditional distributions can not be easily derived or determined from the joint density, and so the Gibbs sampler can not be applied. A most common case in which Gibbs sampling seems to be inappropriate is that the conditional densities are not of known forms for the multi-parameters model. In such situations, a hybrid algorithm may be used. This algorithm tailors the M-H algorithm within a framework of the Gibbs sampler. Such a hybrid algorithm has the following steps to be executed (In the continuation of the algorithm of Gibbs sampler given in the previous section).

1− Start with $j = 1$ and initial values $\{\alpha^{(0)}, \beta^{(0)}\}$.

2– Using initial values $\{\alpha^{(0)},\beta^{(0)}\}$, generate candidate (proposal) points $\{\alpha_c^{(j)},\beta_c^{(j)}\}$ from proposal densities $\{q_1(\alpha^{(j)}|\alpha^{(j-1)}),q_2(\beta^{(j)}|\beta^{(j-1)})\}$, respectively.

3– Calculate the ratios at the point $\alpha_c^{(j)}$ and previous point $\alpha^{(j)}$

$$\rho_1 = \left\{ \frac{\pi_1(\alpha_c^{(j)}|\beta^{(j-1)},x)q_1(\alpha^{(j-1)}|\alpha_c^{(j)})}{\pi_1(\alpha^{(j-1)}|\beta^{(j-1)},x)q_1(\alpha_c^{(j)}|\alpha^{(j-1)})} \right\}$$

4– Accept $\alpha^{(j)}$ as

$$\alpha^{(j)} = \begin{cases} \alpha_c^{(j)} \textit{ with probability } min\ (\rho_1,1) \\ \alpha_c^{(j-1)} \textit{ with probability } 1-min\ (\rho_1,1) \end{cases}$$

5– Then calculate the ratio for β

$$\rho_2 = \left\{ \frac{\pi_2(\beta_c^{(j)}|\alpha^{(j)},x)q_2(\beta^{(j-1)}|\beta_c^{(j)})}{\pi_2(\beta^{(j-1)}|\alpha^{(j)},x)q_2(\beta_c^{(j)}|\beta^{(j-1)})} \right\}$$

6– Accept $\beta^{(j)}$ as

$$\beta^{(j)} = \begin{cases} \beta_c^{(j)} \textit{ with probability } min\ (\rho_2,1) \\ \beta_c^{(j-1)} \textit{ with probability } 1-min\ (\rho_2,1) \end{cases}$$

7– Repeat steps 2 to 6 for all $j = 1,2,\cdots,N$ and obtained $\{(\alpha^{(1)},\beta^{(1)}),(\alpha^{(2)},\beta^{(2)}),\cdots,(\alpha^{(N)},\beta^{(N)})\}$.

1.11 Censoring

The common characteristic of survival data is that one cannot always see the lifetimes of all items under study; rather, for some individuals, the genuine lifetime T is only known to be more or less than some value. Left censored data occurs whenever the censored data points are

below a certain value but the degree by which they are below that value is unknown. Right-censored data, on the other hand, is defined as the set of censored data points that are above a certain value but not known to how much. For illustration, suppose a patient in a clinical trial is moved to another clinic/hospital for treatment and is no longer be eligible for treatment under the same study. The last day on which the patient reported to the clinic for a regular check-up and was known to be alive is the only information about the patient's survival. A patient's actual survival time can also be regarded as censored when death occurs due to a cause that is known to be unrelated to the treatment. Suppose a clinical trial is conducted in order to determine the survival times of some AIDS patients. Other causes of mortality, such as cancer, heart attack, and high blood pressure, may happen in a few patients. The survival times due to some causes other than AIDS are considered censored.

In other circumstances, the experimental items/units may be intentionally removed from the study to reduce the time required for data collecting to manageable levels, as individuals' lifespan can last an indefinite period of time. To reduce the time required to complete the experiment, various types of censoring are included in the design. The concept of censoring is accountable for some of the most significant advancements in survival analysis. For details see the books Lawless (2011), Nelson (2003), Lee and Wang (2003), Klein and Moeschberger (2006) which adduced the techniques for survival data analysis. There are various categories of right censoring, such as time censoring (Type-I censoring), failure censoring (Type-II censoring), and progressive censoring which are described below.

1.11.1 Type-I and Type-II Censoring

Suppose n identical units, here units may be regarded as human beings (patients) or electronic items/systems, are put on a life-test, performed under controlled environment, results in an i.i.d. failure lifetimes. In Type-I censoring, the experiment is terminated at a pre-determined time T_1, and the lifetime of k items failed by this time is observed, and the remaining $n_c = n-k$ items remained alive. The Type-I censored data consists of the lifetimes of k failed items and

the censoring time T_1 for the remaining n_c items. Under this censoring, the time at which the experiment gets terminated is fixed but observable units are random. Type-II censoring allows us to terminate the experiment as soon as a prefixed number (say, $r < n$) of units have failed. Therefore, Type-II censoring ensures the availability of observation on the r failed units for the study but the duration of the life-test is random. For both Type-I and Type-II censoring, the likelihood function is defined by Cohen (1963), Cohen (1965) is given by

$$L(\Theta|x) = r!\binom{n}{r}\prod_{i=1}^{r} f_X(x_i|\Theta)[1-F_X(T_0||\Theta)]^{n-r}, \tag{1.24}$$

where, $f_X(.)$ and $F_X(.)$ are the PDF and CDF respectively, Θ represents the model parameter and may be vector valued, and $\underline{x} = [x_1, x_2, \cdots, x_r]$ denotes the observed data. For Type-II censoring, $T_0 = x_r$, where, x_r denotes the lifetime of the r^{th} item, while for Type-I censoring, T_0 be a pre-determined time for the experiment and r be the number of units have failed by this time. Giving various real life examples with their associated inferences, Lawless (2011) and Nelson (2003) addressed the problem of estimation under survival analysis with Type-I and Type-II censoring. For some more citation one may refer to Sinha et al. (1976), Sinha (1986), Balakrishnan and Aggarwala (2000), Ashour and Afify (2008), Shah and Patel (2011).

1.11.2 Progressive Type-II Censoring with Binomial Removal

A disadvantage of Type-I and Type-II censoring method is that they do not allow the removal of active units during the experiment. Therefore, a general censoring scheme is introduced named as progressive Type-II censoring scheme. This scheme was firstly introduced by Cohen (1963). He suggests that this scheme may be used when test components are very expensive. Progressive censoring allows for both failure (Type-II censoring) and time censoring (Type-I censoring). In the progressive Type-II censoring , m effective sample units are removed at intermediate stage out of n sample units. In this censoring scheme removals are pre-fixed. But, if removals of censoring units are random then this censoring scheme are not suitable method. Then we consider another censoring method that having random removals with progressive

Type-II censoring is known as Progressive Type-II Censoring with Binomial removals (PT-II CBR). In this thesis, PT-II CBR is considered and it can be described as follows.

Let us assume that the experimenter conducts a life test experiment with n items/ units and decides to terminate the experiment as soon as m failure times are recorded. At first failure observed at X_1, R_1 out of the $n-1$ surviving items/ units are randomly removed from the experiment and the experiment continues. Similarly, at second failure observed X_2, R_2 of the remaining $n-R_1-2$ surviving items/ units are again randomly remove from the experiment and in a similar way the experiment continues till the m^{th} failure is recorded and at this stage all the remaining $(n-m-\sum_{i=1}^{m-1} R_i = (R_m))$ surviving items units are removed resulting to termination of the experiment. Since, R_i at i^{th} stage is the total removal out of surviving units, each experiencing the risk of removal with probability p; it is a random variable following the binomial distribution $B(n-m-\sum_{i=1}^{m-1} R_i, p)$. For details see. Viveros and Balakrishnan (1994) and Ng et al. (2004). Following Cohen (1963) for fixed removals, say $R_1 = r_1, R_2 = r_2, R_3 = r_3, \cdots, R_m = r_m$, the conditional likelihood function can be written as,

$$L(\Theta; x|R=r) = c\prod_{i=1}^{m} f(x_i|\Theta)[1-F(x_i|\Theta)]^{r_i}, \cdots -\infty < x_1 < \ldots < x_m < \infty, \tag{1.25}$$

here, Θ is the parameter space, n, m ε N, $1 \leq i \leq m$ and $c = \prod_{i=1}^{m} \gamma_i$ where $\gamma_i = \sum_{j=1}^{m} (r_j + 1)$. Substituting PDF and CDF into (1.25).

It may be noted that type-II censoring is a special case of PT-II CBR with $r = (0, 0, \cdots, 0, n-m)$. For more details on progressive censoring and its further development, readers may be referred to Balakrishnan and Aggarwala (2000), Balakrishnan (2007). There is the massive literature available that reports for estimation of parameters of several lifetime distributions based on progressive censored samples, see Cohen and Norgaard (1977), Davis and Feldstein (1979), Viveros and Balakrishnan (1994), Rastogi and Tripathi (2013), Krishna and Malik (2012), Krishna and Kumar (2013) and Singh et al. (2013b).

For PT-II CBR having recent developments discussed by Yuen and Tse (1996), Tse and Yuen (2000), Yuen and Tse (1996) and Tse et al. (2000).

The assumptions regarding the uniform distribution with equal chance for the number of removals, or binomial distribution with the fixed probability of a removal at each stage, do not seems to be realistic in the practical situations. Consider that a doctor starts an experiment with n patients. The patients may drop from the experiment due to various physical and psychological reasons. For example one of the reasons may be the duration of the cure. Thus at the early stages the chance of drop out will be small as compared to the later stages. The degree of belief, even if not cured completely, may be another factor. If the doctor's cure is not providing immediate relief the chances of drop out at the early stages are expected to be high as compared to the later stages. Hence, keeping these points in mind, it seems more reasonable to think that the number of removals follow a binomial distribution with random probability (p).

1.12 Summary of the Thesis

This thesis has a total of five chapters. Chapter 1 is the introductory definition and terminologies part of the thesis. It contains a brief explanation of the various terms and concepts which have been used in the rest of the thesis. Mainly this thesis work is based on estimation of the parameters for a few lifetime models and ecological model.

Chapter 2, deals with parameter estimation of experimental items/units from the WPD under PT-II CBRs. The EM algorithm has been used for ML estimators. The ML estimators and Bayes estimators have been obtained under symmetric and asymmetric loss functions. The performance of competitive estimators have been studied through their simulated risks. One sample Bayes prediction and expected experiment time have also been studied. Furthermore, through the real bladder cancer data set, the suitability of the considered model and proposed methodology has been illustrated.

The estimators $\hat{\alpha}_G, \hat{\beta}_G$ and $\hat{\lambda}_G$ perform better than all other considered competitive estimators, for ($\delta > 0$, δ is loss parameter) i.e., when o.e. is more serious than u.e. and for ($\delta < 0$) i.e., when u.e. is more serious than o.e., under both considered loss functions. Thus, the use of the proposed estimator $\hat{\alpha}_G, \hat{\beta}_G$ and $\hat{\lambda}_G$ are recommended under SELF and GELF. Moreover, a brief study has done on the expected experiment time by taking the various combinations of effective parameters n, p and m and it observed that on increases the value of p and m, the expected time to test increases. While, for fixed m, on increases the value of n, the expected time to test decreases. The LR test has performed the goodness of fit. The one sample Bayes prediction has also presented. Furthermore, a real data set is fitted to show the practical applicability of the model.

In Chapter 3, we present the E-Bayesian and Bayesian estimators of parameters of PIED under SELF, GELF, and LINEX for PT-II CBRs. The E-Bayesian and Bayesian estimators are compared through risk based on simulated samples. The effectiveness of proposed methodology is applied on the survival time of multiple myeloma patients' data.

The risk of the E-Bayesian and Bayesian estimators of λ and θ are compared under SELF, GELF and LINEX. Generally, we found that the estimated risk of the E-Bayesian estimate of λ and θ have minimum. Therefore, the simulated results shown in this chapter that the E-Bayesian estimation is more efficient and better to perform than Bayesian estimation.

Chapter 4, deals with empirical Bayes estimators of parameter, reliability, and hazard function for Kumaraswamy distribution under the LINEX loss function for PT-II CBRs and Type-II censored samples. The proposed estimators have been compared with the respective Bayes estimators for their simulated risks. The applicability of the proposed estimators has been illustrated through ulcer patient data.

We may conclude that the proposed empirical Bayes estimators $\hat{\lambda}_E, \hat{\lambda}_{E_2}$ and $\hat{h}_E(t), \hat{h}_{E_2}(t)$ are better than Bayes estimators $\hat{\lambda}_B, \hat{\lambda}_{B_2}$ and $\hat{h}_B(t), \hat{h}_{B_2}(t)$ for smaller or larger prior variance ($\sigma = 1, 3$) of β with $a = \pm 1.5$. Also, we have seen that Table 4.1-4.2 under LINEX loss function for the estimators $\hat{R}_E(t)$ & $\hat{R}_{E_2}(t)$ are not always less than those of $\hat{R}_B(t)$ and $\hat{R}_{B_2}(t)$. Since the

risks associated with $\hat{R}_B(t)$ and $\hat{R}_{B_2}(t)$ are smaller than the risk associated with reliability of the empirical estimators. Thus, the use of propose estimator $(\hat{\lambda}_E, \hat{R}_B(t), \hat{h}_E(t))$ and $(\hat{\lambda}_{E_2}, \hat{R}_{B_2}(t), \hat{h}_{E_2}(t))$ under PT-II CBRs and Type-II censoring are used under LINEX loss function respectively.

Chapter 5, deals with a Poisson Lindley distribution as a stochastic abundance model in which the sample is according to the independent Poisson process. We have obtained the maximum likelihood estimators through profile likelihood and the conditional likelihood of the number of species. In the Bayesian estimation of the number of species, we have considered two priors i.e., Jefferey's and reference priors. We obtain the Bayes estimators of the number of species through Jeffery's prior and reference prior. The proposed Bayes estimators have been compared with the corresponding profile and conditional ML estimators for their simulated samples. The Jeffery's and reference priors have considered and compared with the Bayesian approach based on biological data.

The biological data, shows similar results are obtained in estimating the number of species S between the posterior and the maximum likelihood estimators. The asymmetry in both the profile likelihood confidence intervals and the credible intervals accounts for skewed profile likelihoods and skewed posterior distribution. There is a large effect of the model on the estimates, showing a need for more models and a careful model selection technique. The methods of Jeffrey's prior and reference prior give us a way to construct priors that are defined to be non-informative or minimally informative. In the simulated result, we obtain model DIC lesser of Poisson Lindley model with reference prior (PLR) than Poisson Lindley model with Jeffery's prior (PLJ). Therefore, we can propose that Bayes estimate i.e. posterior mean from PLR gives the optimum number of species present there.

The **R** software is used for mathematical computations. This thesis contains a list of references at the end. We realize that an exhaustive list of references related to the problem discussed in the thesis is too big to be reproduced here. Therefore we have included only those references that are cited in the thesis and are directly related to our work.