Chapter 1

Introduction and Basic Terminology

1.1 Introduction

The quality of the products or items is an important task for almost all companies or manufacturers. For example, individual components, objects, products, things, etc. must be monitored and improved by manufacturers. In particular, the lifespan of products is an important quality attribute that manufacturers must examine. The quality of the items or products are directly proportional to their lifespan, which can be monitored by an experimenter using a reliability experiment in which n identical objects or items are put to the test at the same time. One significant problem is that monitoring the failure times of all test units or objects is not always feasible. There are many real-life situations, where the experimenter has to remove some test units or items from the test unintentionally or intentionally due to breakage of test units or time restrictions or findings etc. As a result, censored samples are more appropriate rather than complete samples in life testing experiments. If only $m(< n)$ of the n test units or items placed on life test are detected before the experiment ends, the sample is said to be a censored sample. For modeling such type of data, different censoring schemes are investigated by experimenters in the literature, some of them are as follows:

(a) Type-I or time censoring

(b) Type-II or failure censoring

(c) Random censoring

(d) Hybrid censoring

(e) Progressive censoring

(f) First failure censoring

(g) Progressive first failure censoring.

Out of these censoring schemes, we have employed only the following three censoring schemes in this thesis:

1.1.1 Random Censoring

Random censoring is a natural phenomena of life testing experiments in which the test units or items under study are lost or destroyed before its complete failure. This censoring scheme was first introduced by Gilbert (1962). Type-I censoring is a particular type of random censoring that occurs at a specified time point, say $t = t_0$, see (Lawless, 2003, p. 55). Generally, in survival analysis or clinical trials, this type of censoring commonly occurs in one of the following forms: patients do not complete their treatment and leave before the trial is completed. Random censoring has recently gained popularity in survival analysis and clinical studies. In Chapters 2 and 3 of this thesis, we will go through this censoring scheme in greater depth.

1.1.2 Progressive Censoring

Progressive censoring (PC) is a censoring approach by which test units or items during the life test can be removed or withdrawn from the test at predetermined or random assessment times. Initially in the literature, the progressive Type-II censoring was given by Herd (1956) named as 'multiple censoring'. Later on the PC scheme was inroduced by Cohen (1963) as 'progressive Type-II censoring scheme' in the literature. For more details about PC scheme and their applications one may refer Balakrishnan and Aggarwala (2000) & Balakrishnan and Cramer (2014). The estimation of stress-strength reliability (SSR) based on this censoring scheme in greater depth will be discussed in Chapter 4.

1.1.3 Progressive First Failure Censoring

If the lifespan of goods or objects is very long, then the testing duration of an experiment becomes too long. For such a situation Johnson (1964) introduced the first failure censoring (FFC) scheme and further updated by Balasooriya (1995) that offers a cost-effective and time-saving testing plan for life tests. This censoring scheme allows experimenters to test $k \times n$ test units by testing n groups each containing k test units and then runs all the tests simultaneously until the first failure is observed in each group. Although the FFC scheme is cost-effective and time-saving, it does not enable intermittent removal of units or objects throughout the tests. However, the PC scheme enables removals of units or objects throughout the test. Due to the cost-effectiveness and time-saving properties of the FFC scheme and intermittent removal property of PC scheme, Wu and Kuş (2009) combined them and introduced a more efficient life testing plan called progressive first failure censoring scheme (PFFCS). In Chapter 5 and Chapter 6, we will discuss this censoring scheme in further depth in connection with information theory and survival analysis.

1.2 Estimation Methods

To make any inferences about desire parameters in the statistical theory, estimation methods play important role. In parametric inferences, the mathematical form of the probability distributions are known to us except for a few arbitrary constants associated with the model called as parameters, and our primary concern is to estimate the associated parameters. The classical and Bayesian estimation methods are two popular estimation methods in statistical theory.

1.2.1 Classical Estimation Methods

The classical estimation methods assume the availability of a sample from a specified population, and statistical inference can be developed to have the best long-run performance. In this method, the sample observations are random but the parameters are assumed to be unknown constants. The information about unknown parameters are gathered from the randomness of the sample observations and it is utilized to draw inferences about the unknown parameters. Various classical point estimation methods are widely used in the literature. These methods include the following: method of moments (MM), method of maximum likelihood (ML), method of percentile (MP), method of least square (MLS), method of weighted least square (WLS), method of maximum product spacing's (MPS), method of Anderson-Darling (AD), method of right Anderson-Darling (RAD), method of Cramer-Von-Misses (CVM), etc. Among these methods, the method ML is the most popular and commonly used estimation method in statistical theory. This procedure has numerous advantages and its properties can be utilized in various cases. For more details one may refer (Casella and Berger, 2002, pp. 315-323), (Rohatgi and Saleh, 2015, pp. 388-399). We have employed the ML estimation method in the case of the classical estimation method. Also, we have used the asymptotic confidence interval estimation method for the associated model parameters based on the method of ML. Also, we have employed a well-known resampling technique as the bootstrap method for interval estimations and discussed them in Chapter 5 and Chapter 6.

1.2.2 Bayesian Estimation Method

Bayesian analysis is used in a variety of fields, including science, engineering, medicine, sports, etc. The Bayesian estimation method is based on the prior belief that all the associated model parameters are random variables, allowing previous information to be taken into account. The prior information is used to construct the posterior distribution of the parameter of interest, which is based on the data on lifetimes. This posterior distribution is used to make numerous conclusions about the lifetime parameters and the foundation of Bayes inference.

Suppose that n units are place on a life test and it is assumed that their recorded lifetimes $X_1, X_2, \ldots, X_n$ form a random sample of size n with a population having probability density function (pdf) $f(x|\theta)$, where, θ is a real valued unknown parameter and lies in the parameter space Θ. We also assume that θ is a random variable with pdf $g(\theta)$, which is known as the prior distribution of θ. Thus, the joint pdf of $\{X_1, X_2, \ldots, X_n, \theta\}$ is given by

$$J(\underset{\sim}{x}, \theta) = \prod_{i=1}^{n} f(x_i|\theta)\ g(\theta) = L(\underset{\sim}{x}, \theta) g(\theta), \tag{1.1}$$

where, $L(\underset{\sim}{x}|\theta)$ is the likelihood function. Then, the marginal pdf of $\{X_1, X_2, \ldots, X_n\}$ is given by

$$P(\underset{\sim}{x}) = \int_{\Theta} J(\underset{\sim}{x}, \theta) d\theta. \tag{1.2}$$

Therefore, using the Bayes theorem, the posterior distribution of θ is given by

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

1.2.2.1 Bayesian Approximation Techniques

The posterior mean of any function of the parameters, say $\phi(\theta)$ using posterior distribution in (1.3) is given by

$$E\left[\phi(\theta)|\underset{\sim}{x}\right] = \frac{\int\limits_{\theta\,\in\,\Theta} \phi(\theta) e^{l(\theta;\,\underset{\sim}{x})} g(\theta) d\theta}{\int\limits_{\theta\,\in\,\Theta} \phi(\theta) e^{l(\theta;\,\underset{\sim}{x})} g(\theta) d\theta}, \tag{1.4}$$

where, $l(\theta, \underset{\sim}{x}) = \ln L(\underset{\sim}{x}|\theta)$. From the above posterior mean in equation (1.4), we can say that the posterior mean is in the form of the ratio of two integrals for which the closed-form or solutions may or may not be available. Thus, in general, Bayes estimators are often obtained as a ratio of two integrals for which the closed-form solutions may or may not be available. If closed-form solutions are not available, we require some appropriate approximation techniques to get the approximate Bayes estimators. There are several approximation techniques available in the literature. For example, Lindely and Tierney-Kadane (TK) approximations are used to compute only point estimates, for more details see, Lindley (1980), Sinha (1986) and Tierney and Kadane (1986). Apart from these approximation techniques, a general class of approximation techniques are known as Markov Chain Monte Carlo (MCMC) methods which are used to make inferences based on the posterior samples. These approximation techniques provide both point and interval estimates. Gelfand and Smith (2000), Chen et al. (2000), Robert and Casella (2004), and Gelman et al. (2013) provide comprehensive explanations of MCMC techniques and their applications. In this thesis, we employ TK approximation and MCMC techniques for Bayesian computations.

1.2.2.2 Prior Distributions

A parameter's prior distribution is the probability distribution that expresses the parameter's uncertainty before the current data are observed. There are various forms of priors in the literature, some of them are as follows:

Non-informative Prior: If a prior has very little or no impact on the parameter's posterior distribution, it is said to be non-informative prior.

Informative Prior: If a prior has an effect on the parameter's posterior distribution, it is said to be informative.

Improper Prior: A prior is said to be improper prior if it has infinite density function, i.e. $\int\limits_{\Theta} g(\theta) d\theta = \infty$. For example, $g(\theta) = \frac{1}{c}; \quad \theta > 0$ is an improper prior.

Proper Prior: A prior is said to be proper prior if it has finite probability function, i.e. $\int_{\Theta} g(\theta)d\theta < \infty$. For example $g(\theta) = 1$.

Conjugate Prior: If the prior and posterior distributions belong to the same family of distributions, the prior is said to be a conjugate prior for that family. As a result, the posterior distribution displays the prior distribution in shape.

1.2.2.3 Loss Function

Suppose that the estimator d estimates the unknown parameter θ of pdf $f(x|\theta)$. If the true value of the unknown parameter θ is approximated by d, the loss sustained is indicated by $L(d,\ \theta)$, where $L(d,\theta) = 0$ for $d = \theta$. There are various forms of loss functions in the literature, including squared error loss function (SELF), precautionary loss function (PLF), entropy loss function (ELF), generalised entropy loss function (GELF), and LINEX loss function, among others. Table 1.1 presents the loss functions with their corresponding Bayes estimators.

TABLE 1.1: Loss functions and corresponding Bayes estimators.

Notation	Loss Function	Bayes Estimator
SELF	$\left(\hat{\theta} - \theta\right)^2$	$E(\theta\|\underset{\sim}{x})$
PLF	$\frac{(\hat{\theta}-\theta)^2}{\hat{\theta}}$	$\sqrt{E(\theta^2\|\underset{\sim}{x})}$
ELF	$\left[\frac{\hat{\theta}}{\theta} - \ln\left(\frac{\hat{\theta}}{\theta}\right) - 1\right]$	$\left[E(\theta^{-1}\|\underset{\sim}{x})\right]^{-1}$
GELF	$a\left[\left(\frac{\hat{\theta}}{\theta}\right)^q - q\ln\left(\frac{\hat{\theta}}{\theta}\right) - 1\right]$	$E\left[\theta^{-q}\|\underset{\sim}{x}\right]^{-\frac{1}{q}}$
LINEX	$a\left[e^{b(\hat{\theta}-\theta)} - b(\hat{\theta} - \theta) - 1\right]$	$-\frac{1}{b}\ln\left[E(e^{-b\theta}\|\underset{\sim}{x})\right]$

1.3 Goodness of Fit Test and Model Comparison Criteria

Obtaining information about the population from which a sample is selected is a significant challenge in statistical theory. A statistical model's goodness of fit defines how well it fits a collection of data. The disparity between actual values and predicted values from a model under consideration is often summarised by the goodness of fit measures. In statistical hypothesis testing, such measures can be utilized. Also, the model comparison tests are used to check the performance of the considered model among others. The goodness of fit and model comparison may be done in a variety of ways. The following are some of them:

1.3.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is one of the well known non-parametric goodness-of-fit test. KS test measures the distance between the observed and expected distribution functions. To perform the two-sided goodness of fit test for testing

$$H_0: F(x) = F_0(x) \quad \text{vs} \quad H_1: F(x) \neq F_0(x). \tag{1.5}$$

The KS test statistic is given by

$$D_n = \sup_x \mid F_n(x) - F_0(x) \mid,$$

where, $F_n(x)$ and $F_0(x)$ are the observed and expected distribution functions, respectively. We reject the H_0 if the computed D_n is larger than the critical value, else it may be accepted (see (Conover, 1972, pp. 309–314)).

1.3.2 Anderson-Darling Test

In literature, the Anderson–Darling (AD) goodness-of-fit test was introduced by Anderson and Darling (1954) based on the difference between the observed and expected distribution functions, but here the difference is measured in terms of the square instead of the absolute value used in the KS test. To employ the two-sided AD goodness-of-fit test for testing (1.5). The AD test statistic is given by

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F_0(x_{i:n}) - \ln \bar{F}(x_{n-j+1:n})\right]^2,$$

where, $F_n(x)$ and $F_0(x)$ are the observed and expected distribution functions, respectively. We reject the H_0 if the computed AD statistic A^2 is larger than the critical value, else it may be accepted. For more details one may refer (Gibbons and Chakraborti, 2011, pp. 137–138).

1.3.3 Maximum Likelihood Criterion

The maximum likelihood criterion (MLC) is one of the model selection criteria. It measures the correctness of an estimated statistical model. Based on the specified complete or censored data, different competing models can be ranked according to their maximum likelihood value or equivalently minimal log-likelihood value, with the one having the lowest maximum likelihood value being the best.

1.3.4 Akaike's Information Criterion

The Akaike's information criterion (AIC) was introduced by Akaike (1974). AIC measures the potentiality of estimated statistical models. For a given data set, different statistical models can be ranked according to their AIC, with the one having the lowest AIC being the best. The following formula is used to determine the AIC:

$$\text{AIC} = 2k - 2\log(L), \tag{1.6}$$

where, L is the maximum value of the likelihood function for the estimated model, and k is the number of parameters in the model.

1.3.5 Bayesian Information Criterion

In literature, the Bayesian information criterion (BIC) was proposed by Schwarz (1978) for model selection criteria among a class of statistical models with different number of parameters, say k. Many competing models for a given data set of size n can be ranked according to their BIC, with the model with the lowest BIC being the best, similar to how AIC works. BIC has a greater penalty for extra factors than AIC. The following formula is used to get the BIC:

$$\text{BIC} \ = k\log(n) - 2\log(L) \tag{1.7}$$

1.3.6 Kaplan-Meier Estimator

The product-limit estimator, commonly known as the Kaplan–Meier (KM) estimator, estimates the survival of lifetime data. In literature, the KM estimator was proposed by Kaplan and Meier (1958). A KM survival function estimator curve is a series of horizontal steps with decreasing amplitude that approximate the real survival function for that population. The KM estimator curve has the benefit of being able to handle censored data. The KM estimator curve is identical to the empirical survival function when there is no censoring. For more details one may refer (Lawless, 2003, p. 80).

Several competing models can be rated based on how near the curves of their predicted survival functions are to the KM estimator's curve. The KM estimator is defined as follows:

$$\hat{S}(t) = \prod_{y_i \leq t} \left(1 - \frac{1}{n_i}\right)^{d_i},$$

where, n_i= Number of surviving units at time y_i, and

$$d_i = \begin{cases} 1 & \text{failed/uncensored test units} \\ 0 & \text{censored test units.} \end{cases}$$

1.4 Thesis at a Glance

This thesis consists of six chapters. Chapter 1 is completely introductory in nature with brief discussions of life testing experiments, censoring schemes, classical and Bayesian estimation theories. Also, it presents some useful goodness of fit tests and model selection criteria.

In Chapter 2, we develop classical and Bayesian estimates of the associated model parameters of the randomly censored inverse Pareto (IP) lifetime model. To assess the performance of different estimators, the numerical computations are performed through a Monte Carlo (MC) simulation study. We investigate two randomly censored real data sets based on two different types of cancer illnesses to illustrate the feasibility of the considered model and techniques.

In Chapter 3, we focus on the inverse Weibull (IW) lifetime model with random censoring. The ML and Bayesian estimation approaches for the parameter and reliability characteristics of the IW lifetime model are developed using randomly censored data. The expected time on the test (ETT) is also computed for randomly censored data. Multiple values of real parameters are utilized in the simulation study to investigate the behaviour of these estimators. Finally, a randomly censored real data example is given for demonstration purposes.

Chapter 4 deals with the SSR for the IP lifetime model based on progressively censored data. A system's or machine's SSR is defined as a measure of performance in the context of mechanical durability. The system or machine will fail if the applied stress is greater than the strength of the system or machine at any time point. The ML estimator for SSR and the asymptotic confidence interval (ACI) are developed. The Bayes estimator and HPD credible interval for SSR are obtained using non-informative and informative priors under the GELF. An MC simulation study is used to compare the proposed estimation methods. Finally, two pairs of real data sets are assessed for demonstration reasons.

In Chapter 5, we developed classical and Bayesian inference of associated unknown parameter and reliability characteristics of IP lifetime model based on progressively first failure censored (PFFC) data. For estimation of associated parameter and reliability characteristics, we used ML in the classical estimation process. In addition, asymptotic and bootstrap confidence intervals for the parameter are calculated. The TK approximation, importance sampling, and the Metropolis-Hasting (MH) methods are used to calculate the Bayes estimates of the associated unknown parameter and reliability characteristics. Also, we compute the parameter's HPD credible interval. Extensive numerical computations are conducted to determine the performance of various estimators developed in this chapter. Finally, a real data set is examined to demonstrate the concept and methods suggested.

In Chapter 6, we discussed a problem from information theory based on PFFC data and developed statistical inferences of Shannon's entropy for the Maxwell (MW) lifetime model. Shannon's entropy is an essential quantity that determines the amount of accessible information or the uncertainty of a random process's result. The ML estimates of associated unknown parameter and entropy are computed using the expectation-maximization (EM) algorithm. Also, based on ML estimates we constructed ACIs of parameter and entropy. In addition, we also constructed bootstrap confidence intervals. The Bayes estimators and HPD credible intervals of parameter and entropy are derived under the LINEX loss function. The performance of various estimation methods is compared by an MC simulation study. Finally, real-life data has been analyzed for illustrative purposes.

The statistical software R is used for computations throughout the thesis. Finally, a complete list of references and other literature surveys is given at the end of the thesis as the bibliography. Also, a list of research papers is presented at the end of this thesis.