CHAPTER - I
INTRODUCTION

Applied statisticians frequently face problems dealing with the analysis of time-to-event data. Diverse fields such as medicine, biology, public health, epidemiology, engineering, economics, and demography present such data. Further the analysis of survival experiments is complicated by issues of censoring, where an individual's life length is known to occur only in a certain period of time, and by truncation, where individuals enter the study only if they survive a sufficient length of time or individuals are included in the study only if the event has occurred by a given date. Some events which cause the individual to be randomly censored, with respect to the event of interest, are accidental deaths, migration of human population, death due to some cause other than the one of interest, patient withdrawal from a clinical trial and so forth.

Outliers make statistical analysis difficult. When a sample contains outliers, they may give rise to two superficially distinct problems. Usually, one wishes to use the outliers in the sample to make model inferences that are minimally affected by the number, nature or the values of any observations. Such accommodation methods are covered under the headings of robust and nonparametric inference.

The exponential life distribution provides a good description of the life length of a unit which does not age with time. A property of the exponential distribution which makes it especially important in reliability theory and applications is that the remaining life of a used component is independent of initial age. This property is referred to as "memory less"
property. Other important life distributions are Weibull distribution, Linear Failure Rate distribution and Makeham distribution.

The Weibull distribution has been used to describe fatigue failure (Weibull, 1939), vacuum tube failure (Kao, 1958) and ball bearing failure (Lieblein and Zelen, 1956). It is perhaps the most popular parametric family of failure distributions at the present time. The Weibull distribution belongs to the class of increasing failure rate (IFR) distribution for values of shape parameter $\alpha \geq 1$.

1.1 Organisation of the Thesis:

This dissertation contains 6 chapters. Chapter I is an introductory chapter containing the basic ideas needed for further development.

Chapter II deals with outlier robust estimation of the mean of an exponential distribution under the presence of multiple outliers. Suppose that observations $X_1,\ldots,X_{n-p}$ arise at random from an exponential distribution with density function

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

and observations $X_{n-p+1},\ldots,X_n$ from

$$f(x; \theta \xi) = \frac{\xi}{\theta} e^{-\xi x/\theta} \quad \text{with } \xi < 1.$$  

This is, of course, a scale contamination model for an exponential sample of size $n$ with $p$ contaminants, which might well be expected to reveal themselves as upper outliers. For situation of this type, Kale and Sinha (1971) obtained the estimator for one outlier model. Joshi (1972) extended some of the results of Kale and Sinha (1971) and obtained the values of $m$ for which $T_m$ has the smallest mean squared error (MSE).
For further references on this, one can refer Sinha (1973), Kale (1975), Patel, Mudholkar and Fernando (1988) and the review discussion in Barnett and Lewis (1993).

A particular case of some interest is the full sample L-estimator of Chikkagoudar and Kunchur (1980) in the class

\[ \tilde{\theta} = \sum_{i=1}^{n} \alpha_i X_{(i)} \]  

where

\[ \alpha_i = \frac{1 - \frac{2i}{n(n+1)}}{n} \]

where \(X(1) \leq X(2) \leq \ldots \leq X(n)\) are the order statistics of the sample. This can be interpreted as a linear combination

\[ \tilde{\theta} = \sum_{i=1}^{n} \beta_i \tilde{\theta}_i \]

of separate estimators

\[ \tilde{\theta}_i = \bar{X} - \frac{X_{(i)}}{n} \]

where \(\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i\), with weight

\[ \beta_i = \frac{\frac{2i}{n(n+1)}}{ \frac{2i}{n(n+1)}} = 1 \]

This estimator in (1.1) has been shown to present varying degrees of relative efficiency when compared with alternative estimators proposed by Kale and Sinha (1971, a winsorized mean) and Joshi (1972), and the (0,1) trimmed mean. It is sometimes more efficient, sometimes less so.

The principal characteristic of the estimator (1.1) is that the more extreme upper sample values are down weighted—but only moderately so.
The weights do not decrease rapidly enough to provide adequate compensation against contaminants, particularly if there are several of these and they are manifest as upper outliers.

Balakrishnan and Barnett (2002) proposed a modification of the Chikkagoudar and Kunchur estimator with weights that decrease more rapidly at the upper end of the sample. The generalised estimator of Balakrishnan and Barnett is given below

\[ \tilde{\theta}_k = \frac{1}{n} \sum_{i=1}^{n} \left[ 1 - \frac{(k + 1)(i + k - 1)^{(k)}}{(n + k)^{(k+1)}} \right] X_{(i)} \]

where \((i+k-1)^{(k)} = (i+k-1)(i+k-2)...i\).

This class includes Chikkagoudar and Kunchur estimator as a special case corresponding to \(k=1\). For comparative purpose, the estimator

\[ T_n = \frac{1}{n+1} \sum_{i=1}^{n} X_{(i)} = \frac{\bar{X}}{n+1} \]

is also included. In this chapter we propose estimators based on the probabilities of different order statistics being outliers in the \(p\) outlier model. Based on these probabilities bias and mean squared error of the proposed estimators are obtained for different values of \(n\), \(p\) and \(\xi\). A paper containing the results of this chapter has been presented at National Conference on Statistical Inference and its applications, held at Shivaji University, Kolhapur and sent for publication.

Increasing failure rate average (IFRA) distributions naturally arise when coherent systems are formed from components with independent increasing failure rate distributions. Also it naturally arises when one considers cumulative damage shock models. The class of increasing failure rate average distributions contains the class of increasing failure
rate distributions as subclass. IFRA class of distributions is the smallest class of distributions which contains the exponential distribution and is closed under formation of coherent systems.

Let $F$ be a probability distribution such that $F(0)=0$. Then $F$ is an increasing failure rate average distribution (IFRA) if $\left[\frac{F(t)}{t}\right]$ is decreasing in $t > 0$, or equivalently, for $x > 0$, $0 < b < 1$,

$$F(bx) \geq \{F(x)\}^b$$

where $\bar{F} = 1 - F$. The equality in (1.2) holds if and only if $F$ is an exponential distribution.

Tests for exponentiality designed to detect the alternative hypotheses relevant in reliability theory include those of Proschan and Pyke(1967), Bickel and Doksum(1969), Ahmed(1975), Hollander and Proschan(1972, 1975) and Koul(1977, 78). However, the test that have been developed specifically for testing for increasing failure rate average alternative only is due to Deshpande(1983), besides tests due to El-Bassiouny (2003) and Hendi and Abouammoh (2001).

The generalization of the class of tests for exponentiality against IFRA alternatives is considered in Chapter III. In this chapter, we propose a new class of test statistic which are U-statistics whose kernel depends on sub sample minima. An alternative expression based on ranks for the proposed class of statistics is also given. The asymptotic normality and consistency of the proposed class of tests is studied. Pitman ARE comparisons, unbiasedness of the tests and empirical powers of the tests are studied. A paper based on this chapter has been sent for publication.

Testing exponentiality against nonparametric aging life classes has been active area of research over the past three decades. One of the
important classes tested is the class of 'increasing failure rate' (IFR). Let $X$ be a nonnegative random variable with distribution function $F$ and survival function $\bar{F} = 1 - F$. We say that $X$ is IFR if $\frac{\bar{F}(x + y)}{\bar{F}(x)}$ is decreasing in $x \geq 0$ for all $y \geq 0$. This definition is equivalent to saying that the failure rate $r(x) = \frac{f(x)}{\bar{F}(x)}$ is nondecreasing provided that the probability density function $f(x)$ of $F(x)$ exists. Many authors have considered the problem of testing the hypothesis that $F$ is exponential against the alternative that it is IFR. The early contributors are Proschan and Pyke (1967), Barlow and Proschan (1969), Bickel and Docksum (1969) and Bickel (1969) to mention a few.

A major problem concerning the procedures suggested for testing against the IFR alternative is that these procedures are not easy enough to introduce them into undergraduate or beginning graduate nonparametric statistics courses or texts. This is mainly because calculating these test statistics is often too complicated to be done by hand and may require extensive programming on computer. Even when the statistics are easily calculable, their limiting distributions are not easy to derive. To remedy this situation, Ahmad (2004) provided a new simple approach and proposed tests for testing exponentiality against IFR and NBU alternatives. In Chapter IV, we propose a new class of simple test statistics based on U-statistics for testing exponentiality against IFR alternatives. Its properties and performance in terms of Pitman ARE are investigated. A paper based on the results of this paper has been submitted for publication.

Mathisen (1943) proposed a test for the two-sample location problem based on the number of observations in the $X$-sample not
exceeding the median of the Y-sample. Besides Mathisen’s test, a number of distribution free tests are available in the literature. Wilcoxon rank sum test is a popular nonparametric procedure for this problem. Mood’s median test is particularly effective in detecting shift in location in distributions which are symmetric and heavy tailed. The normal scores test is detecting shift in a normal distribution. Gastwirth’s L and H tests are effective in detecting shifts in moderately heavy tailed distributions. The RS test due to Hogg, Fisher and Randles (1975) is effective in detecting shifts in distributions that are skewed. The test proposed by Shetty and Govindarajulu(1988), Shetty and Bhat(1993) based on subsample medians takes care of specified number of outliers at the extremes of both the samples. The tests proposed by Shetty and Bhat(1994) are resistant to outliers in one of the samples.

Chapter V deals with the effect of random censoring on the performance of a class of tests proposed by Shetty and Bhat (1994). ARE comparisons are carried out for various distributions.

Stochastic dependence between random measurements is one of the important aspects of many statistical investigations. The quantification of this concept for bivariate distributions has been attempted by several authors. One of the oldest measures based on ranks is Spearman’s rank correlation coefficient. Some other well-known measures of concordance are Kendall’s tau, Blomquist’s Q and Hoeffding’s $\Delta$ to mention a few.

Suppose $(X,Y)$ is an absolutely continuous random variable with joint distribution function $F(x,y)$ and survival function $\bar{F}(x,y)$. We
consider the problem of testing the null hypothesis of independence against positive quadrant dependence. That is

\[ H_0 : F(x, y) = F_1(x) \cdot F_2(y) \forall x, y \]

against

\[ H_1 : F(x, y) \geq F_1(x) \cdot F_2(y) \text{ with strict inequality holding on a set of nonzero probability} \]

where \( F_1(x) \) and \( F_2(y) \) are marginal distribution functions of \( X \) and \( Y \) respectively.

In Chapter VI, we consider Kendall’s sample tau coefficient for testing \( H_0 \) against \( H_1 \) when the variable \( X \) is subjected to random censoring. The mean, variance and asymptotic distribution of Kendall’s sample tau coefficient in this setup are obtained in Section 3. In Section 4, Pitman ARE of Kendall’s tau in the censored setup relative to uncensored setup is obtained. Various censoring distributions are investigated for Morgenstern and Woodworth family of distributions.