Chapter 1

Introduction, Preliminaries and Summary

1.1. Introduction

There is a plethora of examples of phenomena concerning nature, life and human activities where the real data do not conform to the standard distributions. In such cases, we either use mixtures of standard distributions of similar types or non-standard mixtures of degenerate distribution and a standard distribution which may be again a discrete or continuous one. Pearson (1894) was possibly the first to study formally the case of a mixture of two normal distributions. The literature contains many papers that deal with mixtures of distributions of similar types, such as the mixture of normal distributions, the mixture of chi-square distributions, the mixture of exponential distributions, and the mixture of binomial distributions and so on (Robbins and Pitman, 1949). However, the literature contains very few papers that provide and deal with special "nonstandard" mixture that mix discrete (degenerate, even) and continuous distributions as emphasized in this thesis. In general, the word mixture refers to a convex combination of distributions or random variables. We consider a case of a two-component univariate complete mixture model where one component's distribution function (DF) is F_1 and the other component's DF is F_2 while the mixing proportion p unknown. Such a model can be defined in its most general form as

$$G = (1 - p)F_1 + p\,F_2, 0 \leq p \leq 1. \tag{1.1.1}$$

The model in (1.1.1) is called a mixture of distribution functions F_1 and F_2. In terms of random variables, one may say that G models an observation Z that are obtained as follows: With probability $(1-p)$ observe X having distribution F_1 and with probability p observe Y having distribution F_2. Such a mixture may then be viewed as a model for data that may be interpreted as the outcomes of a two-stage experiment. In the first stage, a population is randomly chosen and then in the second stage, an observation is made from the chosen population. It is not necessary to limit oneself to mixtures of just two or even a finite number of distributions. It should also be emphasized that there is considerable ambiguity associated with the mixture in infinitely many ways. Nevertheless, when mixture models are formulated reasonably, they can provide useful tools for statistical analysis. There is by now a large literature pertaining to statistical analysis of mixtures of distributions (Titterington, et al., 1985). Problems and applications of mixtures also appear in the literature associated with the term heterogeneity (Keyfitz, 1984).

Non-standard mixture of distributions generally contains inliers and outliers. Inlier is an observation (or a group of observations) sufficiently small relative to the rest of the observations, which appears to be inconsistent with the remaining dataset. Kale and Muralidharan (2000) have introduced the term *inliers* in connection with the estimation of (p, θ) of early failure model with modified failure time distribution (FTD) as an exponential with mean θ assuming p known. Inliers are either the result of the instantaneous failures or the early failures or both, experienced in life testing experiments, clinical trials, weather predictions, geographic information systems, athlete performance analysis, and many other such applications. The test items that fail at time 0 are called the *instantaneous failures* and the test items that fail prematurely are called the *early failures*. These occurrences may be due to the inferior quality of a product or service, or faulty construction or alignment of events/objects, or due to no response to the treatments. Such failures usually discard the assumption of a single mode distribution and hence the usual method of modeling and inference procedures may not be accurate in practice. These situations need special attention, as the modified model is a nonstandard mixture distribution. That is a model in which, with probability $(1-p)$, a specified constant is observed whereas, with probability p, one observes a random measurement whose distribution has a density function. That is, it is a mixture of degenerate distribution and an absolutely continuous one.

The problem of outliers is well-known in statistics: an outlier is a value that is far from the general distribution of the other observed values, and can often perturb the results of a statistical analysis. Various procedures exist for identifying and studying outliers (Barnett and Lewis, 1994). An inlier, by contrast, is an observation lying within the general distribution of other observed values, generally does not perturb the results but is nevertheless non-conforming and unusual. For single variables, an inlier is practically impossible to identify, but in the multivariate case, thanks to interrelationships between variables, values can be identified that are observed to be more central in a distribution but would be expected, based on the other information in the data matrix, to be more outlying. In that sense, the lower outlier may be treated as inliers.

Upon a thorough literature survey, we have found some vague definitions of inliers. Some authors have defined inliers as those observations which are not outliers. While outliers are erroneous observations located farther away from the sample mean, inliers are erroneous observations located closer to the mean, (Akkaya and Tiku, 2005). According to UN publication (UNECE, 2000), an inlier is a data observation that lies in the interior of a dataset and is unusual or an error. Because inliers are difficult to distinguish from the other data values, they are sometimes difficult to find and – if they are in error – to correct.

Aitchison (1955) was the first to discuss the inference problem of instantaneous failures in life testing. The author has provided the efficient estimation of parametric functions under various probability models. Some earlier studies on these type of models are treated by Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991, 1995), Kale (1998), and Shinde and Shanubhougue (2000), Dixit (2003). The inferences on inliers was studied in detail by Kale and Muralidharan

(2000, 2008), Muralidharan and Lathika (2004, 2006), Muralidharan (2010), Muralidharan and Arti (2008, 2013), Bavagosai and Muralidharan (2016, 2018) and Muralidharan and Bavagosai (2016a,b, 2017, 2018).

1.2. Examples of inliers

There are many practical contexts, where inliers can be natural occurrences of the specific situations involved and degeneracy can happen at one or more discrete points and a positive distribution for the remaining lifetimes. Some of the situations are as follows:

1. In an audit sample, we have two pieces of information, namely, the booking amount (recorded) and the audited amount (correct). The difference between the two is called the error amount. Here some population elements contain no error, whereas other population elements contain error of varying amounts. The distribution of errors can, therefore, be viewed as two distinguishable distributions, one with a discrete probability mass at ‘zero’ (no error) and the other a continuous distribution of non-zero positive and/or negative error amounts. The data here can be modeled using a nonstandard mixture.
2. In the mass production of technological components of hardware, intended to function over a period of time, the failure rate is initially relatively high, and then actually decreases with increasing age. The high failure rate either results in zero life time or marginally small life times, otherwise, the life time will be any positive number. Thus, the overall distribution of lifetimes may be represented by a nonstandard mixture.
3. Time until remission is of interest in studies of drug effectiveness for treatment of certain diseases. Some patients respond and some do not respond to the treatment. This is an example of a distribution having a mixture of the mass point at 0 which corresponds to instantaneous remission and a nontrivial continuous distribution having positive remission times.

4. In a study of tooth decay, the numbers of surfaces in a mouth which are filled, missing or decayed are scored to produce a decay index. Healthy teeth are scored 0 for no evidence of decay. Thus, the distribution is a mixture of a mass point at 0 and a nontrivial continuous distribution of decay score. The problem could be further complicated if the decay score is expressed as a percentage of damage to measured teeth. The distribution should then be a mixture of a discrete random variable (0-for healthy teeth, 1-for all teeth missing) with a nonzero probability of both outcomes and a continuous random variable (amount of decay in the $(0, 1)$ interval).
5. In the studies of genetic birth defects, children can be characterized by two variables, a discrete variable to indicate if one is affected and born dead and a continuous variable measuring the survival time of affected children born alive. We may consider here a mixture of the mass point at 0 and a nontrivial continuous distribution of survival time of affected children.
6. In studies of methods for removing certain behaviors (e.g., predatory, behavior or salt consumption), the amount of the behavior which is exhibited at a certain point in time may be measured. In this context, the complete absence of the target behavior may represent a different result than would a reduction from a baseline level of the behavior. Thus, one would model the distribution of activity levels as a mixture of a discrete value of zero and a continuous random level.
7. The number of plants per plot affected by a disease may be represented by Poisson family but the observation at 'zero' i.e. unaffected plants may be due to the Poisson probability at 'zero' and $(1-p)\%$ of the plants were resistant to the disease due to other causes. This is a nonstandard indistinguishable mixture.
8. The first response time of patients during a medical operation: Consider measurements of physical performance scores of patients with a debilitating disease such as multiple sclerosis. There will be frequent zero measurements from those giving no performance and many observations with graded positive performance.

9. In measuring precipitation amounts for a specified time period, one must deal with the problem that a proportion of these amounts will be zero (i.e. measured as zero). The remaining proportion is characterized by some positive random variable. The distribution of this positive random variable usually looks reasonable smooth, but in fact is itself a complex mixture arising from many different types of events.
10. In studying the human smoking behavior, two variables of interest are smoking status- ‘ever smoked’ and ‘never smoked’ and a score on a ‘pharmacological scale’ of people who have smoked. Here the data is modeled with two discrete value points, namely 0 (for never smoked), 1 (for ever smoked) and a “pharmacological score” in continuous measurement. A nontrivial conditional distribution of the second variate can be defined only in association with the first outcome of the first variate.
11. In the Study of tumor characteristics, two variates may be recorded. The first is the absence (0) or presence (1) of a tumor and the second is tumor size measured on a continuous scale. In this problem, it is sometimes of interest to consider a marginal tumor measurement that is 0 with nonzero probability and therefore leading to a case of mixture of unrelated distributions.
12. Machines and software’s are tested for its correctness and perfectness or reliability. Bugs in such situation are important to assess the durability and credibility of machines and programs. Zero defects or zero bugs are considered to be good in such situations. If there are bugs, then it can be measured in terms of some discrete measurements.
13. In a community, a particular service, such as a specific medical care, may not be utilized by all families in the community. There may be a substantial portion of non-takers of such a service. Those families who subscribe to it do so in varying amounts. Thus, the distribution of consumption of service may be represented by a mixture of zeros and positive values. Thus, the overall distribution of consumption of service may be represented by a nonstandard distinguishable mixture.

14. In a quite different context, important problems exist in time series analysis in which there are mixed spectra containing both discrete and continuous components.

As it is obvious that, there are many such practical situations, where data can evolve with a mix of discrete and continuous measurements. Hence, probability modeling in such situations may require special attention and treatment. One may also see ‘Statistical models and analysis in Auditing: Panel on nonstandard mixtures of distributions’, appeared in *Statistical science* (1989) journal for similar such examples and the details. This thesis addresses the inference procedures for models of nonstandard distributions, based on some of the above examples.

1.3. Inlier prone models

From the above examples, it is seen that the values including zeros and close to zeros are important as well as significant in most cases. For instance, zero errors in an audit report, zero tooth decay, and zero bugs in a computer program or electronic machine are all good to judge the prevailing situation, and hence they are significant. Similarly, zero lifetime, zero rainfall (dry day) etc., are all situationally bad but again significant as per the conditions prevailing. Thus inliers are more natural than outliers, where most of the time inliers are retained after the detection and considered for future analysis. As a consequence, the modeling of inliers distribution is more important than its detection for statistical decision making. Below, we introduce various inliers prone models, and then delve into the theoretical treatment of these inliers models in various chapters. Many practical problems discussed above are modeled, studied and applied by suggesting fit of distributions, estimation of parameters and testing of hypothesis etc.

1.3.1. Instantaneous failure model

Since instantaneous failures are a natural occurrence and such failures usually discard the assumption of a unimodal distribution and hence traditional way of modeling and inference procedures may not be accurate in practice. To tackle these situations the model is represented as

$$G(x;p,\theta) = \begin{cases} 1-p, & x=0 \\ 1-p+p\,F(x;\theta), & x>0 \end{cases} \tag{1.3.1}$$

with respect to a measure μ which is the sum of Lebesgue measure on $(0,\infty)$ and a singular unit measure at the origin; and $0<p<1$. Here the model $\mathcal{F} = \{F(x;\theta), x \geq 0, \theta \in \Theta\}$ where $F(x;\theta)$ is a continuous failure time distribution function with $F(0)=0$ is to be suitably modified as a non-standard mixture of distribution by mixing a singular distribution at zero to accommodate instantaneous failures. After Aitchison (1955), the other authors who have studied this model are Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991, 1995), Kale (1998, 2003), Muralidharan (1999), Muralidharan and Kale (2002), Muralidharan and Lathika (2005, 2006), Kale and Muralidharan (2006, 2007, 2008), Adlouni et al. (2011) and so on. In this thesis, authors have considered other commonly used parametric models such as Gompertz, Weibull, Pareto, Lindley distributions.

1.3.2. Early failure model-1

If it is assumed that $\lambda(x) = \lambda = \frac{1}{\theta}$ for all x from an exponential distribution, then the assumption of an exponential density is equivalent to the assumption of a constant failure rate. Under this setup, Miller (1960), proposes the early failure model as

$$\lambda(x) = \begin{cases} \lambda_1, & 0 \leq x < T_0 \\ \lambda_2, & T_0 \leq x \end{cases} \tag{1.3.2}$$

where $\lambda_1 > \lambda_2$. The probability density correspond to this failure rate is

$$f_X(x;\lambda_1,\lambda_2) = \begin{cases} \lambda_1 e^{-\lambda_1 x}, & 0 \le x < T_0 \\ \lambda_2 e^{-\lambda_1 T_0 - \lambda_2 (x - T_0)}, & T_0 \le x \end{cases} \tag{1.3.3}$$

The justification follows from the fact that when a component is put on the test, it is not known whether it is an 'early failure' or a 'standard' item. Since some will be early failures, the failure rate on the average will be high at the start, but if an item has survived for a certain period of time T_0, then it cannot be an 'early failure' so its failure rate will be lower for the succeeding time period. The above model can also be viewed as a model for a shift in the hazard function of exponential distribution. Further analysis of this models are presented in Chapter 2.

1.3.3. Early failure model-2

To accommodate early failures, the family $\mathcal{F}$ is modified to $\mathcal{G}_1 = \{G(x;p,\theta), x \ge 0, \theta \in \Theta, 0 < p < 1\}$, where the cumulative distribution function (CDF) corresponding to $g_1 \in \mathcal{G}_1$ is given by

$$G_1(x;p,\theta) = (1-p)\,H(x) + p\,F(x;\theta). \tag{1.3.4}$$

Here $H(x)$ is a CDF with $H(\delta) = 1$ for δ sufficiently small, assumed known and specified in advance. Then the modified family $\mathcal{G}_1$ has a probability density function (pdf) with respect to measure μ, which is the sum of Lebesgue measure on (δ, ∞) and a singular measure at δ as

$$g_1(x;p,\theta) = \begin{cases} 0, & x < \delta \\ 1 - p + p\,F(\delta;\theta), & x = \delta \\ p\,f(x;\theta), & x > \delta \end{cases} \tag{1.3.5}$$

Some of the references which treats early failure analysis with exponential distributions are Kale and Muralidharan (2000, 2007, 2008), Kale (2003), Muralidharan and Lathika (2006), Muralidharan and Arti (2008, 2013), Muralidharan (2010), Muralidharan and Bavagosai (2016a) and the references contained therein. These authors treated early failures as inliers using the sample configurations from other parametric models including Weibull, Pareto, Normal and Gompertz distributions. The models in (1.3.1) and (1.3.4) can be combined to form the CDF as

$$G(x;p,\theta)=\begin{cases}0, & x<d\\ (1-p)+pF(x;\theta), & x\geq d\end{cases} \tag{1.3.6}$$

with the corresponding probability density function (pdf) as:

$$g(x;p,\theta)=\begin{cases}0, & x<d\\ 1-p+pF(d;\theta), & x=d\\ pf(x;\theta), & x>d\end{cases} \tag{1.3.7}$$

If $d=0$ the model reduces to the instantaneous failures case and if $d>0$, it reduces to the case of early failures. In fact, the models given in (1.3.5) and (1.3.7) are same. One may also see Lai et al. (2007) for a complete mixture model, where they have treated the instantaneous part based on Dirac delta function and a probability distribution for the positive observations.

1.3.4. Model with inliers at zero and one

In some of the examples discussed above (e.g. 4, 5, 10, 11), the observations 0 and 1 become a natural occurrence with other positive observations. If these observations are treated as inliers, then, the distribution function of such models can be written as

$$H(x;p_1,p_2,\theta)=\begin{cases}0, & x<0\\ p_1, & 0\le x<1\\ p_1+p_2, & x=1\\ p_1+p_2+(1-p_1-p_2)\frac{F(x;\theta)-F(1;\theta)}{1-F(1;\theta)}, & x\ge 1\end{cases} \tag{1.3.8}$$

where p_1 and p_2 are the proportion of 0 and 1 observations. This model was first studied by Muralidharan and Bavagosai (2017, 2018), Bavagosai and Muralidharan (2018) with $F(x;\theta)$ as exponential, Weibull and Pareto. One can also use other probability models for $F(x;\theta)$. In various chapters of this thesis we study these models. In Chapter 8, we discuss a generalization of the above model.

1.4. Preliminaries

To set the flow of the thesis, we need to use some existing literatures on statistical terms and theorems. They are mostly based on the book by Lehman and Casella (1986), and are presented below:

Distinguishable mixture: The mixtures are distinguishable in the sense that one can tell from which population an observation has come, otherwise the mixtures are called indistinguishable.

Unbiased estimator: An estimator $\delta(x)$ of $g(\theta)$ is unbiased if

$$E_\theta[\delta(x)]=g(\theta) \text{ for all } \theta\in\Omega.$$

Exponential family: A family $\{P_\theta\}$ of distributions is said to form an s-dimensional exponential family if the distributions P_θ have densities of the form

$$p_\theta(x)=exp\left[\sum_{i=1}^{s}\eta_i(\theta)T_i(x)-B(\theta)\right]h(x) \tag{1.4.1}$$

with respect to some common measure μ. Here, the η_i and B are real-valued functions of the parameters, $h \geq 0$ and the T_i are real-valued statistics, and x is a point in the sample space $\mathfrak{x}$, the *support* of the density.

Sufficient statistics: A statistics T is said to be sufficient for X, or for the family $\mathcal{P} = \{P_\theta,\ \theta \in \Omega\}$ of the possible distribution of X, or for θ, if the conditional distribution of X, given $T = t$ is independent of θ for all t .

Theorem 1.4.1. (Factorization criterion). *A necessary and sufficient condition for a statistic T to be sufficient for a family $\mathcal{P} = \{P_\theta \in \Omega\}$ of distributions of X dominated by a σ–finite measure μ is that there exist non-negative functions g_θ and h such that the densities p_θ of P_θ satisfy*

$$P_\theta(x) = g_\theta[T(x)]h(x) \qquad (a.\ e.\ \mu).$$

Complete family: (V. K. Rohatgi, 1976). Let $\{f_\theta(x), \theta \in \Omega\}$ be a family of pdf's (or pmf's). We say that this family is complete if

$$E_\theta[g(x)] = 0 \text{ for all } \theta \in \Omega \text{ implies } P_\theta[g(x) = 0] = 1 \text{ for all } \theta \in \Omega.$$

Complete statistic: A statistic distributions of $T(x)$ is said to be complete if the family of distributions of T is complete.

Theorem 1.4.2. *If X is distributed according to the exponential family (1.4.1) and the family is of full rank, then $T = [T_1(x)], \dots, [T_s(x)]$ is complete.*

*Remark***:** Suppose $X = (X_1, \dots, X_n)$ has a distribution $\mathcal{P} = \{P_\theta \in \Omega\}$, belonging to the one parameter exponential family. Then the statistic $T(x)$ is called the natural sufficient statistic for the family $\{P_\theta\}$.

Minimal sufficient statistics: A sufficient statistic T is said to be *minimal* if of all sufficient statistics it provides the greatest possible reduction of the data, that is, if for any sufficient statistic U there exists a function H such that $T = H(U)$ (a.e. P).

Corollary 1.4.1. **(Minimal sufficient statistics for exponential families)**. Let X be distributed with density (1.4.1), then $T = (T_1, \dots, T_s)$ is minimal sufficient provided the family (1.4.1) satisfies one of the following conditions:

(i) It is of full rank.

(ii) The parameter space contains $s + 1$ points $\eta^{(j)}(j = 0, \dots, s)$, which span E_s in the sense that they do not belong to a proper affine subspace of E_s.

Note: Since complete sufficient statistics are particularly effective in reducing the data, it is not surprising that a complete sufficient statistic is always minimal. Proofs are given in Lehmann and Scheff´e (1950), Bahadur (1957), and Schervish (1995).

Uniformly Minimum Variance Unbiased Estimator: An unbiased estimator $\delta(x)$ of $g(\theta)$ is the uniformly minimum variance unbiased estimator (UMVUE) of $g(\theta)$ if $Var_\theta\ \delta(x) \le Var_\theta\ \delta'(x)$ for all $\theta \in \Omega$ where $\delta'(x)$ is any other unbiased estimator of $g(\theta)$.

The existence, uniqueness, and characterization of UMVUEs have been investigated by Barankin (1949) and Stein (1950).The relationship of unbiased estimators of $g(\theta)$ with unbiased estimators of zero can be helpful in characterizing and determining UMVUEs when they exist.

Theorem 1.4.3. *Let X have distribution $P_\theta, \theta \in \Omega$, let δ be an estimator in Δ, and let $\mathcal{U}$ denote the set of all unbiased estimators of zero which are in Δ. Then, a necessary and sufficient condition for δ to be a UMVUE of its expectation $g(\theta)$ is that*

$$E_\theta(\delta U) = 0\ \forall\ U \in \mathcal{U}\ \ and\ \theta \in \Omega.$$

Likelihood function: For a sample point $X = (X_1, X_2, \dots, X_n)$ from a density $f(\boldsymbol{x};\theta)$, the likelihood function $L(\theta|x) = f(\boldsymbol{x};\theta)$ is the sample density considered as a function of θ for fixed $\boldsymbol{x}$. In the case of iid observations, the likelihood is $L(\theta|\underline{x}) = \prod_{i=1}^{n} f(x_i;\theta)$. It is then often easier to work with the logarithm of the likelihood function, $l(\theta|\underline{x}) = \sum_{i=1}^{n} log\, f(x_i;\theta)$.

Information matrix: Suppose $X_1, X_2, \dots, X_n$ are an iid random sample of size n with pdf $f(x;\ \theta)$. Then the quantity $E_\theta\left\{\left[\frac{\partial \log f(x;\,\theta)}{\partial\theta}\right]^2\right\}$ is called the information number or Fisher information of the sample. This terminology reflects the fact that the information number gives a bound on the variance of the best-unbiased estimator of θ. As the information number gets bigger and we have more information about θ, we have a smaller bound on the variance of the best-unbiased estimator. The information matrix contains the variance and covariance of the elements of the score vector. Thus the generic element of the information matrix, in the $(i,j)^{th}$ position, is $E\left[-\frac{\partial^2 \log L(\theta|\underline{x})}{\partial\theta_i\,\partial\theta_j}\right] = E\left[\frac{\partial \log L(\theta)}{\partial\theta_i}\ \frac{\partial \log L(\theta|\underline{x})}{\partial\theta_j}\right]$.

Type II censored scheme: (Lawless, 2003). A Type II censored sample is one for which only the r smallest observations in a sample of size n are observed, $1 \le r \le n$ where r is determined before the data are collected. Let n lifetimes of the sample be $X_1, X_2, \dots, X_n$. Their order statistics are $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. In Type II censoring we know only the values $X_{(1)}, X_{(2)}, \dots, X_{(r)}$. Let f$(x)$ be the pdf of the lifetimes, f$(x)\,dx$ = probability of end-of-life $X \in [x, x+dx]$, then the joint pdf of $X_{(1)}, X_{(2)}, \dots, X_{(r)}$ is

$$f_n\big(X_{(1)}, X_{(2)}, \dots, X_{(r)}\big) = \frac{n!}{(n-r)!} f\big(X_{(1)}\big) f\big(X_{(2)}\big) \dots f\big(X_{(r)}\big)\big[S\big(X_{(r)}\big)\big]^{n-r} \tag{1.4.2}$$

where, $S(x) = P(X \ge x) = \int_x^\infty f(u)\,du$ is the survival function.

Masking effect: (Barnett and Lewis, 1994). For k actual inliers, and r hypothesized inliers, with $r < k$, a first inlier mask a second if the second inlier is only identified as an inlier when the first is not present. That is, considering $r < k$ inlier, $k - r$ inliers have been left in the sample and may skew the statistics enough so that the r hypothesized inliers do not appear very extreme. The larger the $k - r$ inliers remaining in the sample, worse the masking.

Swamping effect: (Barnett and Lewis, 1994). For k actual inliers, and r hypothesized inliers, with $r > k$, an inlier swamps a non-inlier when the non-inlier is only identified as an inlier when considered in presence of the inlier. That is, when $r - k$ non-inliers are grouped with k inliers within a test statistic, the test may still reject the null hypothesis, especially if the inliers are small.

The model selection criteria: There are various statistics of fit to indicate how well the estimated model fits the data. The statistics belong to two categories: likelihood based statistics and the empirical distribution function (EDF) based statistics. The statistics: Akaike information criterion (AIC), and Bayesian information criterion (BIC) are likelihood based statistics and statistics Kolmogorov-Smirnov (K-S), Cramer-von Mises (CVM), and Anderson-Darling (AD) are EDF based statistics widely used in model selection criteria.

AIC: The Akaike information criterion is named after the statistician Hirotugu Akaike, who formulated it in (1973). When a statistical model is used to represent the process that generated the data, the model will almost never be exact; so some information will be lost by using the model to represent the process. AIC estimates the relative information lost by a given model: the less information a model loses, the higher the quality of that model. Suppose that we have a statistical model of some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is $AIC = 2k -$

$2ln(\hat{L})$. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

BIC: The Bayesian information criterion is proposed by Gideon E. Schwarz (1978) and hence also referred to as the Schwarz information criterion. and Schwarz Bayesian information criterion, is another model selection criterion based on information theory but set within a Bayesian context. The difference between the BIC and the AIC is the greater penalty imposed for the number of parameters by the former than the latter. The BIC is computed as $BIC = k\ ln(n) - 2ln(\hat{L})$, where *n* is the number of recorded measurements, *k* is the number of estimated parameters, and $\hat{L}$ be the maximum value of the likelihood function for the mode. Given a set of candidate models, the best model is the one that provides the minimum BIC.

K-S test: (Kolmogorov–Smirnov, 1933). Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be an ordered random sample and the distribution of X is $F(x)$. The EDF $F_n(x)$is defined as the fraction of X_i's that are less than or equal to x for each x. That is $F_n(x) = \frac{number\ of\ observations \leq x}{n}$, $-\infty < x < \infty$. The KS statistic belongs to the supremum class of EDF statistics and this class of statistics is based on the largest vertical difference between the hypothesized and empirical distribution. This test requires that the null distribution $F^*(x)$ be completely specified with known parameters. In KS test of normality, $F^*(x)$ is taken to be a normal distribution with known mean μ and standard deviation σ. The test statistics is defined differently for the following three different set of hypotheses. For a right-tailed test $H_0: F(x) = F^*(x)$ versus $H_1: F(x) > F^*(x)$, the test statistic $KS^+ = sup[F^*(x) - F_n(x)]$ is the greatest vertical distance. Likewise, for the left-tailed test $H_0: F(x) = F^*(x)$ versus $H_1: F(x) < F^*(x)$, the test statistic $KS^- = sup[F_n(x) - F^*(x)]$ is the greatest vertical distance. The Kolmogorov–Statistic for a two-sided test, $H_0: F(x) = F^*(x)$ versus$H_0: F(x) \neq F^*(x)$, is taken to be $KS = \max(KS^-, KS^+)$.

CVM test: (Cramer–von Mises, 1928). The CVM test judges the goodness of fit of a hypothesized distribution $F^*(x)$ compared with the EDF $F_{(n)}(x)$ based on the statistic defined as

$$nw^2 = n\int_{-\infty}^{\infty}\left[F_{(n)}(x) - F^*(x)\right]^2 dF(x)$$

which, like the KS statistic, is distribution-free, i.e. its distribution does not depend on the hypothesized distribution, $F^*(x)$. The CVM test is an alternative to the KS test.

AD test: (Anderson–Darling, 1954). The AD test is actually a modification of the CVM test. It differs from the CVM test in such a way that it gives more weight to the tails of the distribution than does the CVM test. Unlike the CVM test which is distribution-free, the AD test makes use of the specific hypothesized distribution when calculating its critical values. Therefore, this test is more sensitive in comparison with the CVM test. A drawback of this test is that the critical values have to be calculated for each specified distribution. The AD test statistic

$$AD = n\int_{-\infty}^{\infty}\frac{\left[F_{(n)}(x)-F^*(x)\right]^2}{F_{(n)}(x)-F^*(x)}\,dF(x).$$

1.5. Estimation procedures

Below, we discuss those estimation procedures used in this thesis.

1.5.1. Maximum Likelihood Estimation

The method of maximum likelihood is, by far, the most popular technique for computing estimators. The principle of maximum likelihood is relatively straightforward. The principle of maximum likelihood yields a choice of the estimator $\hat{\theta}$ as the value for the parameter that makes the observed data most probable. If $X_1, X_2, \dots, X_n$ are n an iid sample from a population with pdf (or pmf) $f(x; \underline{\theta})$, where $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_s)$ then the likelihood function is

$$L(\underline{\theta}|\underline{x}) = \prod_{i=1}^{n} f(x_i; \underline{\theta})$$

The maximum likelihood estimator (MLE) of the parameters $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_s)$ is $\underline{\hat{\theta}}_{MLE} = (\hat{\theta}_{1MLE}, \hat{\theta}_{2MLE}, \dots, \hat{\theta}_{sMLE})$, say, obtained by maximizing $L(\underline{\theta}|\underline{x})$. That is, $\underline{\hat{\theta}}_{MLE}$ is obtained by solving

$$\frac{\partial L(\underline{\theta}|\underline{x})}{\partial \theta_i} = 0, i = 1,2, \dots, s. \tag{1.5.1}$$

It is then often easier to work with the logarithm of the likelihood function, and $\underline{\hat{\theta}}_{MLE}$ obtained by solving

$$\frac{\partial \log L(\underline{\theta}|\underline{x})}{\partial \theta_i} = 0, i = 1,2, \dots, s. \tag{1.5.2}$$

1.5.2. Least Squares and Weighted Least Squares Estimation

They are regression-based estimators of the unknown parameters, which were originally suggested by Swain et al. (1988) to estimate the parameters of Beta distributions. The method can be described as follows: Suppose $X_1, X_2, \dots, X_n$ is a random sample of size n from a distribution function $F(x; \underline{\theta})$ and $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denotes the order statistics of the observed sample. It is well known that

$$E\left[F\left(X_{(i)};\underline{\theta}\right)\right]=\frac{i}{n+1} \text{ and } V\left[F\left(X_{(i)};\underline{\theta}\right)\right]=\frac{i(n-i+1)}{(n+1)^2(n+2)}$$

See, Johnson et al. (1995) for details and properties. Using the expectations, two variants of the least square methods can be proposed as follows:

Least Squares Estimators: The least squares estimator (LSE) of the parameters $\underline{\theta}=(\theta_1,\theta_2,\dots,\theta_s)$ is $\underline{\hat{\theta}}_{LSE}=\left(\hat{\theta}_{1LSE},\hat{\theta}_{2LSE},\dots,\hat{\theta}_{sLSE}\right)$, say, can be obtained by minimizing

$$\sum_{i=1}^{n}\left[F\left(X_{(i)};\underline{\theta}\right)-\frac{i}{n+1}\right]^2 \tag{1.5.3}$$

with respect to the unknown parameters $\theta_i, i=1,2,\dots,s$.

***Weighted Least Squares Estimators*:** The weighted least squares estimator (WLSE) of the parameters $\underline{\theta}=(\theta_1,\theta_2,\dots,\theta_s)$ is $\underline{\hat{\theta}}_{WLSE}=\left(\hat{\theta}_{1WLSE},\hat{\theta}_{2WLSE},\dots,\hat{\theta}_{sWLSE}\right)$ can be obtained by minimizing

$$\sum_{i=1}^{n}w_i\left[F\left(X_{(i)};\underline{\theta}\right)-\frac{i}{n+1}\right]^2 \tag{1.5.4}$$

with respect to the unknown parameters $\theta_i, i=1,2,\dots,s$,

where $w_i=\frac{1}{v\left(F\left(X_{(i)};\underline{\theta}\right)\right)}=\frac{(n+1)^2(n+2)}{i(n-i+1)}$.

1.5.3. Percentile Estimation

This method was originally explored by Kao (1958, 1959). If the data comes from a distribution function which has a closed form, then it is quite natural to estimate the unknown parameters by fitting a straight line to the theoretical points obtained from the distribution function and the sample percentile points. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from a distribution function $F\left(x; \underline{\theta}\right)$ and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denotes the order statistics of the observed sample. If P_i denotes some estimate of $F\left(x; \underline{\theta}\right)$, then the percentile estimator (PE) of the parameter $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_s)$ is $\underline{\hat{\theta}}_{PE} = \left(\hat{\theta}_{1PE}, \hat{\theta}_{2PE}, \ldots, \hat{\theta}_{sPE}\right)$, say, can be obtained by minimizing

$$\sum_{i=1}^{n}\left[\log P_i - \log F(x_{(i)}; \underline{\theta})\right]^2 \qquad (1.5.5)$$

with respect to $\theta_i, i = 1,2, \ldots, s$. Several estimators of P_i can be used here (see, Murthy et al., 2004). In this thesis, we mainly consider $P_i = \frac{i}{n+1}$, which is the expected value of $F\left(X_{(i)}; \underline{\theta}\right)$.

1.5.4. Maximum Product Spacings Estimation

Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from a univariate distribution function $F\left(x; \underline{\theta}\right)$ with corresponding pdf $f\left(x; \underline{\theta}\right)$ and it is required to estimate $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_s)$. The density is assumed to be strictly positive in an interval (a, b) and zero elsewhere, $a = -\infty$ and $b = +\infty$ may also be taken. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the i^{th} order statistics. The maximum product spacings method choose that value of parameter θ which provides the maximum for product spacings

$$S = \prod_{i=1}^{n+1} D_i \qquad (1.5.6)$$

or the average spacing,

$$G = S^{1/(n+1)} \qquad (1.5.7)$$

where the spacing's of the sample is $D_i = F\left(x_{(i)}; \underline{\theta}\right) - F\left(x_{(i-1)}; \underline{\theta}\right), i = 2,3, \dots, n$, $\sum D_i = 1$, $F\left(x_{(0)}; \underline{\theta}\right) = 0$ and $F\left(x_{(n+1)}; \underline{\theta}\right) = 1$.

If there are ties in the data, anticipated difficulty may arise in drawing inferences based on the product of spacings. To get rid of the problem of ties, the method suggested by Shao and Nahan (1999) may be followed: Suppose that among n observations $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ there are m distinct values expressed in ascending order of their magnitudes as $y_{(j)}; j = 1,2, \dots m$. Here, $F\left(y_{(0)}; \underline{\theta}\right) = 0$ and $F\left(y_{(m+1)}; \underline{\theta}\right) = 1$. Let I_j denote the number of observations in $(y_{(j-1)}, y_{(j)}\,]$. Naturally, $\sum_{j=1}^{m} I_j = n$. In light of the above, and assuming $I_{m+1} = 1$, the product spacings in presence of ties, is to choose that value of parameter θ which provides the maximum for S, where

$$S = \prod_{j=1}^{m+1} \left[\frac{F\left(y_{(j)}; \underline{\theta}\right) - F\left(y_{(j-1)}; \underline{\theta}\right)}{I_j}\right]^{I_j} \tag{1.5.8}$$

The other way to tackle the tie problem is to consider that all the equal observations are in fact unequal but differ by the amount smaller than the least count of measurement and hence noted as equal. Suppose two observations x and y are equal then we may consider that actually, the observations as x and $x + dx$. Hence such tied observation should contribute to product of spacings equal to $\lim_{dx \to 0} [F(x + dx) - F(x)]$ which can be approximated by $f(x)$, where $f(.)$ denotes the density function corresponding to F. Thus the modified product spacings to be maximized can be written as

$$S = \prod_{j=1}^{m+1} \left[F\left(y_{(j)};\, \underline{\theta}\right) - F\left(y_{(j-1)};\, \underline{\theta}\right)\right]\left[f\left(y_{(j)};\, \underline{\theta}\right)\right]^{(I_j - 1)}.$$

1.5.5. UMVU Estimation

The uniformly minimum unbiased estimation facilitates the estimation of parameters and parametric functions including probability density function and survival function of a given probability model. The method also facilitates the computation of variances of the estimators in models given in (1.3.6) and (1.3.8). The problem of UMVU estimation can be addressed through two approaches: the traditional approach is by considering the conditional expectation of an unbiased estimator given complete sufficient statistic. For the proposed model studied in this thesis, obtaining conditional distribution given the sufficient statistics is bit difficult. For exponential distribution case, it is possible to obtain the conditional distribution and hence UMVU estimation of some parametric functions is possible (See Chapter 2, Section 2.2 for details).The other approach is by using the method given by Roy and Mitra (1957), where the model is suitably expressed in the form of an exponential family, and thereby use the sufficient statistics and its properties to obtain the UMV estimators. See also Khatri (1959), Patil (1963a,c), Joshi and Park (1974), Charalambides (1974), Jani (1977), Gupta (1977), Patel (1978), Jani and Dave (1990), Jani (1993), Muralidharan (2000), and Singh (2007).

Consider the models incorporating inliers as given in (1.3.6). If $F(x;\theta)$ belongs to the class of one-parameter exponential family of distributions, then the form of probability density function can be written as $f(x;\theta) = a(x)\frac{[h(\theta)]^{d(x)}}{g(\theta)}, x > d, \theta \in \Theta$ where $a(x) > 0$, $d(x)$ is a monotone increasing function of x, $g(\theta)$ and $h(\theta)$ are differentiable functions of θ and $g(\theta) = \int_{x>d} a(x)[h(\theta)]^{d(x)}\, dx$. Then, the probability function of the mixture family (1.3.7) is obtained as

$$g_X(x;\, p, \theta) = (1-p)^{I(x)} p^{[1-I(x)]} \left[\frac{a(x)[h(\theta)]^{d(x)}}{g(\theta)}\right]^{[1-I(x)]}$$

$$= \frac{[a(x)]^{[1-I(x)]}[h(\theta)]^{[1-I(x)]d(x)}\left[\frac{g(\theta)(1-p)}{p}\right]^{[1-I(x)]}}{\frac{g(\theta)}{p}} \tag{1.5.9}$$

where $I(x)$ is an indicator function such that

$$I(x) = \begin{cases} 1, & if\ x = d \\ 0, & o.w. \end{cases} \tag{1.5.10}$$

The mixture density (1.5.9) so obtained, is a well-known form of a two-parameter exponential family with natural parameters $(\eta_1, \eta_2) = \left(\log\left(g(\theta)\ \frac{(1-P)}{p}\right),\ \log\left(h(\theta)\right)\right)$ generated by underlying indexing parameters (p, θ). Hence $\left(I(x), \left(1 - I(x)\right)d(x)\right)$ is jointly minimal sufficient for (p, θ), as $I(x)$ and $\left(1 - I(x)\right)d(x)$ do not satisfy any linear restriction (See Kale and Muralidharan, 2015). The $\eta's$ too do not satisfy any linear constraint and hence the natural parameter space is convex set in E_2 containing a two-dimensional rectangle making (1.5.9) a full rank family. The statistic $\left(I(x), \left(1 - I(x)\right)d(x)\right)$ is thus complete.

Let $X_1, X_2, \dots, X_n$ be a random sample from (1.5.9). Then the joint pdf of the sample is given by

$$g_X\left(\underline{x}; p, \theta\right) = \binom{n}{r} p^r (1-p)^{n-r} \frac{1}{\binom{n}{r}} \prod_{x_j > d} a\left(x_j\right) \frac{[h(\theta)]^z}{[g(\theta)]^r} \tag{1.5.11}$$

$$= \prod_{x_j > d} a\left(x_j\right) \frac{[h(\theta)]^z \left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n} \tag{1.5.12}$$

where $z = \sum_{j=1}^{n}\left[1 - I\left(x_j\right)\right]d\left(x_j\right)$, $n - r = \sum_{j=1}^{n} I\left(x_j\right)$, r is the number of observations greater than d and $d(x) = d$ for $x = d$. By Neyman factorization theorem, $(n - R, Z)$ is sufficient for (p, θ). The joint density of the sample is the product of the distribution of R and the conditional distribution of the sample given

$R = r$ assumes the form of the well-known exponential family. Hence $(Z|r)$ or $(Z|n-r)$ is complete sufficient for θ.

It is known that under random sampling the exponential structure is preserved by the distribution of the sufficient statistics (refer Lehmann and Casella, 1998). Hence, without loss of generality, the conditional pdf of $(Z|n-r)$ can be written as

$$g(z;\ \theta|n-r) = \begin{cases} \dfrac{B(Z|n-r)[h(\theta)]^z}{[g(\theta)]^r}, z > d;\ 0 \le n-r < n \\ 1, \qquad\qquad\qquad\quad z = d;\ r = 0 \end{cases}$$

which depends only on θ and is a complete family of distributions. Here $B(Z|n-r)$ is a function of Z and r, and can be identified suitably for a given model. The distribution of $n-R$ is binomial which is the same as that of R and its probability mass function (pmf) is given by

$$P(R = r) = \binom{n}{r}\ p^r(1-p)^{n-r}, r = 0,1, \dots, n.$$

This also is a complete family. Jayade (1993) has shown that the joint distribution of a random variable (X_1, X_2) containing two parameters θ_1 and θ_2 is complete if the marginal distribution of X_1 is discrete which depends only on θ_1 and belongs to a complete family whereas the conditional distribution of X_2 given X_1 depends only on θ_2 and belongs to a complete family of distributions. Hence $(n-R, Z)$ is complete sufficient for (p, θ). The joint distribution of $(n-R, Z)$ therefore was given by

$$g_Z(n-r, z; p, \theta) = P(n-R = n-r)g(z; \theta|n-r) = P(R = r)g(z; \theta|r).$$

That is,

$$g_Z(z, r; \theta, p) = \begin{cases} \dfrac{B(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}, z > d;\ r = 1,2, \dots, n \\ (1-p)^n, \qquad\qquad\qquad\quad z = d;\ r = 0 \end{cases} \tag{1.5.13}$$

where

$$B(z,r,n)=\begin{cases}\binom{n}{r}B(z|r), & z>d;\ r=1,2,\dots,n\\ 1, & z=d;\ r=0\end{cases} \tag{1.5.14}$$

is such that

$$(1-p)^n+\sum_{r=1}^{n}\int_{z>d}B(z,r,n)\frac{[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz=1.$$

Let $\phi(p,\theta)$ be a function of the parameter p and θ in $g_X(x;p,\theta)$ in (1.5.9). Then for the construction of UMVUE of $\phi(p,\theta)$ on the basis of a random sample from (1.5.9), it is sufficient to find out an unbiased estimator of $\phi(p,\theta)$ which is a function of the complete sufficient statistics $(n-R,Z)$. We now use Roy and Mitra (1957) technique, on the line similar to that of Jani and Singh (1995), to find such an unbiased estimator. The following lemma provides the necessary and sufficient condition for the existence of UMVUE of $\phi(p,\theta)$.

Lemma 1.5.1. *Let $X_1,X_2,\dots,X_n$ be a random sample from (1.5.9). Then there exists UMVUE of $\phi(p,\theta)$ if and only if $\phi(p,\theta)$ can be expressed in the form*

$$\phi(p,\theta)=\alpha(d,0,n)(1-p)^n+\sum_{r=1}^{n}\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz$$

Proof: To prove the *sufficient* part, let

$$\psi(Z,R,n)=\frac{\alpha(Z,R,n)}{B(Z,R,n)},B(Z,R,n)\neq 0.$$

Then,

$$E[\psi(Z,R,n)] = \alpha(d,0,n)(1-p)^n$$

$$+\sum_{r=1}^{n}\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz$$

$$= \phi(p,\theta).$$

Since $\psi(Z,R,n)$ is a function of complete sufficient statistic $(n-R,Z)$, it is the UMVUE of $\phi(p,\theta)$.

The *necessary* part can be proved as follows: Suppose $\psi(Z,R,n)$ is an unbiased estimator of $\phi(p,\theta)$. Then

$$\psi(d,0,n)(1-p)^n+\sum_{r=1}^{n}\int_{z>d}\frac{\psi(z,r,n)B(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz = \phi(p,\theta).$$

That is,

$$\alpha(d,0,n)(1-p)^n+\sum_{r=1}^{n}\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz = \phi(p,\theta)$$

where

$$\alpha(Z,R,n) = \psi(Z,R,n)B(Z,R,n).$$

This completes the proof. ■

Thus, the UMVUE of a function $\phi(p,\theta)$ of θ and p in $g_X(x;p,\theta)$ is given by

$$\psi(Z,R,n) = \frac{\alpha(Z,R,n)}{B(Z,R,n)}, B(Z,R,n) \neq 0, \tag{1.5.15}$$

as a direct consequence of the above lemma.

Lemma 1.5.2. *The UMVUE of a function of θ alone does not exists.*

Proof: Let $\phi(p,\theta) = \varphi(\theta)$, say, a function of θ only and $\psi(Z,R,n)$ be its unbiased estimator based on the complete sufficient statistic $(n-R,Z)$. Hence

$$E\big(\psi(Z,R,n)\big) = \psi(d,0,n)(1-p)^n + \sum_{r=1}^{n}\int_{z>d}\psi(z,r,n)\frac{B(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz = \varphi(\theta).$$

That is,

$$\alpha(d,0,n)(1-p)^n + \sum_{r=1}^{n}\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z\left[\frac{(1-p)g(\theta)}{p}\right]^{n-r}}{\left[\frac{g(\theta)}{p}\right]^n}dz = \varphi(\theta).$$

That is,

$$\alpha(d,0,n) + \sum_{r=1}^{n}\left(\frac{p}{1-p}\right)^r\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z}{[g(\theta)]^r}dz = \varphi(\theta)(1-p)^{-n}.$$

Substituting $\tau(p) = \frac{p}{1-p}$ in the above we get

$$\alpha(d,0,n) + \sum_{r=1}^{n}\tau^r(p)\int_{z>d}\frac{\alpha(z,r,n)[h(\theta)]^z}{[g(\theta)]^r}dz = \varphi(\theta)\big(1+\tau(p)\big)^n.$$

That is,

$$\alpha(d,0,n)+\sum_{r=1}^{n}\tau^r(p)\int_{z>0}\frac{\alpha(z,r,n)[h(\theta)]^z}{[g(\theta)]^r}dz=\ \varphi(\theta)\left(1+\sum_{r=1}^{n}\binom{n}{r}\tau^r(p)\right).$$

By comparing the coefficient of $\tau^r(p)$ on both the sides, we get

$$\alpha(d,0,n)=\varphi(\theta),$$

which is contradicting in nature. Hence a parametric function of θ alone is not U-estimable. ■

Although, the parametric function of θ alone is not U-estimable but with the help of Lemma 1.5.1, we can still estimate other parametric functions like $(1-p)$, the probability density function, survival function etc. in the following result. The proof of results for $d=0$ is given in Singh (2007).

Result 1.5.1. *Suppose $X_1,X_2,\dots,X_n$ is a random sample from (1.5.9). Then, for $m\leq n$, the UMVUE of $(1-p)^m$ (the m^{th} power of the probability of $x=d$) is given by*

$$G_m(z,r,n)=\frac{B(z,r,n-m)}{B(z,r,n)}=\begin{cases}\frac{\binom{n-m}{r}}{\binom{n}{r}}, r=0,1,\dots,n-m\\ 0, \qquad o.w.\end{cases} \tag{1.5.16}$$

Result 1.5.2. *The UMVUE of the variance of $G_m(Z,R,n)$ for $m\leq\frac{n}{2}$ is given by*

$$\widehat{var}[G_m(z,r,n)]=\begin{cases}G_m^2(z,r,n)-G_{2m}(z,r,n), r=1,2,\dots,(n-2m)\\ G_m^2(z,r,n), \qquad r=(n-2m+1),\dots,(n-m)\\ 0, \qquad otherwise\end{cases}$$

$$= \begin{cases} \left[\dfrac{\binom{n-m}{r}}{\binom{n}{r}}\right]^2 - \dfrac{\binom{n-2m}{r}}{\binom{n}{r}}, & r = 1,2,\dots,(n-2m) \\ \left[\dfrac{\binom{n-m}{r}}{\binom{n}{r}}\right]^2, & r = (n-2m+1),\dots,(n-m) \\ 0, & o.w. \end{cases} \tag{1.5.17}$$

Result 1.5.3. *Suppose $X_1, X_2, \dots, X_n$ is a random sample from (1.5.9). Then, for $k > 0$,*

$$H_k(z,r,n) = \frac{B(z,\ r-[1-I(z)]k,\ n-k)}{B(z,\ r,\ n)} \tag{1.5.18}$$

is the UMVUE of the $(1-p)^n + \left[\frac{p}{g(\theta)}\right]^k [1-(1-p)^{n-k}]$.

Result 1.5.4. *The UMVUE of the variance of $H_k(Z,R,n)$ is given by*

$$\widehat{var}[H_k(z,r,n)] = H_k^2(z,r,n) - H_{2k}(z,r,n). \tag{1.5.19}$$

Result 1.5.5. *Let $X_1, X_2, \dots, X_n, n > 1$ be a random sample from (1.5.9). Then for fixed x,*

$$\phi_x(z,r,n) = \begin{cases} a(x)\frac{B(z-d(x),r-1,n-1)}{B(z,r,n)}, & x > d; z > d(x); r = 1,2,\dots,n \\ \frac{B(z,r,n-1)}{B(z,r,n)}, & x = d;\ r = 0,1,\dots,n-1 \\ 0, & o.w. \end{cases} \tag{1.5.20}$$

is the UMVUE of the pdf $g_X(x;p,\theta)$.

Result 1.5.6. *The UMVUE of the variance of $\phi_x(Z,R\ ,n)$ is given by*

$$\widehat{var}[\phi_x(z,r,n)]$$
$$= \begin{cases} \phi_x^2(z,r,n) - \phi_x(z,r,n)\phi_x(z-d(x),r-1,n-1), & \\ & x > d;\ z > 2x;\ r = 2,3,\dots,n \\ \phi_x^2(z,r,n), & x > d; x < z < 2x \\ \phi_x^2(z,r,n) - \phi_x(z,r,n)\phi_x(z,r,n-1), & x = d;\ r = 0,1,\dots,n-1 \\ 0, & o.w. \end{cases} \tag{1.5.21}$$

*Corollary 1.5.1***.** For a fixed z and r, the UMVUE of the Survival function $S(t) = P(X > t), t \geq 0$, is then given by

$$\hat{S}(t) = \int_{x>t} \phi_x(z,r,n)\, dx. \tag{1.5.22}$$

Result 1.5.7. *For fixed z and r, the UMVUE of the $var\left[\hat{S}(t)\right]$ is given by*

$$\widehat{var}\left[\hat{S}(t)\right]$$

$$= \begin{cases} \hat{S}^2(t) - 2\int_{x>t}\int_{y>t} a(x)a(y)\dfrac{B(z-d(x)-d(y),r-2,n-2)}{B(z,r,n)}\, dx\, dy, & \\ & z > 2d(t);\ r = 3,4,\dots,n \\ \hat{S}^2(t), & d(t) < z \leq 2d(t);\ r = 2,3,\dots,n \\ 0, & o.w. \end{cases} \tag{1.5.23}$$

The Minimum Variance Unbiased Estimation in a multi-parameter exponential family of distributions also exploits the concepts of complete sufficient statistics. Consider the r-parameter exponential family of distributions defined by the pdf

$$f(x;\underline{\theta}) = a(x)b(\underline{\theta})e^{\sum_{j=1}^{r}\eta_j(\underline{\theta})c_j(x)},$$

$x \in T \subseteq \mathbb{R}$, $\underline{\theta} = (\theta_1,\theta_2,\dots,\theta_r) \in \Omega$. The above density can also be written as

$$f(x;\underline{\theta}) = a(x)\prod_{j=1}^{r}\frac{[h_j(\underline{\theta})]^{C_j(x)}}{g(\underline{\theta})} \tag{1.5.24}$$

where $a(x) > 0$, $C_j(X)$, $j = 1, \dots, r$ are nontrivial real-valued statistics, $g(\underline{\theta})$ and $h_j(\underline{\theta})$ are at least twice differentiable functions of $\underline{\theta}_j$, $j = 1, \dots, r$ and $g(\underline{\theta}) = \int a(x)\prod_{j=1}^{r}[h_j(\underline{\theta})]^{C_j(x)}\,dx$. Here X may be r-vector, in which case $X \in T(r) \subseteq \mathbb{R}^r$. It can be seen that $C(X) = (C_1(X), C_2(X), \dots, C_r(X))$ is jointly complete sufficient and the distribution of $C(X)$ is also the r-parameter exponential family. Let us study the distributional properties of $C(X)$ first.

Distributional properties of $\boldsymbol{C(X)}$

Since the moments of $C(X)$ are functions of $\underline{\theta}$, they are MVUE's of these functions. Hence, in the following, we shall find the moments of $C(X)$ first. Differentiating $g(\underline{\theta})$ partially with respect to $\theta_1, \theta_2, \dots, \theta_r$ under the regularity conditions, we get

$$\begin{aligned}
\frac{\partial \log g}{\partial \theta_1} &= E(C_1(X))\frac{\partial \log h_1}{\partial \theta_1} + E(C_2(X))\frac{\partial \log h_2}{\partial \theta_1} + \cdots + E(C_r(X))\frac{\partial \log h_r}{\partial \theta_1} \\
\frac{\partial \log g}{\partial \theta_2} &= E(C_1(X))\frac{\partial \log h_1}{\partial \theta_2} + E(C_2(X))\frac{\partial \log h_2}{\partial \theta_2} + \cdots + E(C_r(X))\frac{\partial \log h_r}{\partial \theta_2} \\
&\vdots \\
\frac{\partial \log g}{\partial \theta_r} &= E(C_1(X))\frac{\partial \log h_1}{\partial \theta_r} + E(C_2(X))\frac{\partial \log h_2}{\partial \theta_r} + \cdots + E(C_r(X))\frac{\partial \log h_r}{\partial \theta_r}
\end{aligned} \tag{1.5.25}$$

which can be rewritten as

$$\underline{G} = A\,\underline{\mu},\ |A| \neq 0 \tag{1.5.26}$$

where

$$\underline{G} = \left[\frac{\partial \log g(\underline{\theta})}{\partial \theta_i}\right]_{r\times 1}, i = 1,2,\dots,r$$

$$\underline{\mu} = \left[E\big(C_i(X)\big)\right]_{r\times 1}, i = 1,2,\dots,r$$

$$A = \left[\frac{\partial\, log\, h_i}{\partial \theta_i}\right]_{r\times r}, i = 1,2,\dots,r \text{ and } j = 1,2,\dots,r.$$

According to Cramer's rule, the solution to (1.5.26) is

$$E\big(C_i(X)\big) = \frac{|A_i|}{|A|}, i = 1, 2, \dots, r$$

where A_i is obtained by replacing the i[th] column of A by the elements of $\underline{G}$. The joint moments of $C_1^{k_1}(X),\ C_2^{k_2}(X), \dots,\ C_r^{k_r}(X)$ are given as

$$E\left(C_1^{k_1}(X)\ C_2^{k_2}(X) \dots C_r^{k_r}(X)\right)$$
$$= \int_{x\in T} C_1^{k_1}(x)\ C_2^{k_2}(x) \dots C_r^{k_r}(x)\ a(x) \prod_{j=1}^{r} \frac{\left[h_j(\underline{\theta})\right]^{C_j(x)}}{g(\underline{\theta})} dx$$

which on differentiating with respect to $\theta_1, \theta_2, \dots, \theta_r$ and using (1.5.26), give a system of r linear non-homogeneous equations

$$\underline{G}_1 = A\, \Sigma,\ |A| \neq 0 \tag{1.5.27}$$

where

$$\underline{G}_1 = \begin{bmatrix} \frac{\partial}{\partial\theta_1} E\left(C_1^{k_1}(X)\ ,\ \ \dots\ ,\ \ C_r^{k_r}(X)\right) \\ \frac{\partial}{\partial\theta_2} E\left(C_1^{k_1}(X)\ ,\ \ \dots\ ,\ \ C_r^{k_r}(X)\right) \\ \vdots \\ \frac{\partial}{\partial\theta_r} E\left(C_1^{k_1}(X)\ ,\ \ \dots\ ,\ \ C_r^{k_r}(X)\right) \end{bmatrix}$$

$$\Sigma = \Big[E\left(C_1^{k_1}(X)\ ,\dots\ ,C_i^{k_i+1}(X),\dots,C_r^{k_r}(x)\right) -$$
$$E\big(C_i(X)\big)E\left(C_1^{k_1}(X)\ ,\dots\ ,C_i^{k_i}(X),\dots,C_r^{k_r}(X)\right)\Big]_{r\times r} = [\sigma_{i(1,2,\dots,r)}]_{r\times r}, \text{ (say).}$$

Using Cramer's rule for the solution of a system of linear equations (1.5.27) gives

$$\sigma_{i(1,2,\dots,r)} = \frac{|A_i|}{|A|}\ , i = 1,2,\dots,r \tag{1.5.28}$$

where A_i is obtained by replacing the i^th column of A by the column vector $\underline{G}_1$. For $k_i = 1$ and $k_j = 0\ \forall\ i \neq j = 1,2,\dots,\ r$, we get covariance between $C_i(X)$ and $C_j(X)$ as

$$\sigma_{ij} = \frac{|A_i\ |_{(k_i=1;k_j=0),i\neq j}}{|A|} \tag{1.5.29}$$

Thus, we have the variance-covariance matrix Σ as

$$\Sigma = \big[\sigma_{ij}\big]_{r\times r} = \frac{\big(|A_i\ |_{(k_i=1;k_j=0),i\neq j}\big)}{|A|} \tag{1.5.30}$$

If A_{ij} is the cofactor of the element a_{ij} of A, then

$$|A_i\ |_{(k_i=1;k_j=0),i\neq j=1,2,\dots,r} = \sum_{j=1}^{r} A_{ji}\frac{\partial}{\partial\theta_j}E\big(C_i(X)\big)$$

and hence,

$$\Sigma = \frac{1}{|A|}\begin{bmatrix} \sum_{j=1}^{r} A_{j1}\frac{\partial}{\partial\theta_j}E\big(C_1(X)\big) & \sum_{j=1}^{r} A_{j1}\frac{\partial}{\partial\theta_j}E\big(C_2(X)\big) & \cdots & \sum_{j=1}^{r} A_{j1}\frac{\partial}{\partial\theta_j}E\big(C_r(X)\big) \\ \sum_{j=1}^{r} A_{j2}\frac{\partial}{\partial\theta_j}E\big(C_1(X)\big) & \sum_{j=1}^{r} A_{j2}\frac{\partial}{\partial\theta_j}E\big(C_2(X)\big) & \cdots & \sum_{j=1}^{r} A_{j2}\frac{\partial}{\partial\theta_j}E\big(C_r(X)\big) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{r} A_{jr}\frac{\partial}{\partial\theta_j}E\big(C_1(X)\big) & \sum_{j=1}^{r} A_{jr}\frac{\partial}{\partial\theta_j}E\big(C_2(X)\big) & \cdots & \sum_{j=1}^{r} A_{jr}\frac{\partial}{\partial\theta_j}E\big(C_r(X)\big) \end{bmatrix} \tag{1.5.31}$$

Now, we study the UMVU Estimation in r-parameter exponential family of distributions: Let $X_1, X_2, \dots, X_n$ be a random sample from (1.5.24). It can be easily verified that for exponential family of distributions given by (1.5.24), $Z = (Z_1, Z_2, \dots, Z_r)$, where $Z_i = \sum_{j=1}^{n} C_i(X_j)\,, i = 1,2, \dots, r$ are jointly complete sufficient. The joint pdf of Z is again exponential family and is given by

$$f_Z\big(z; \underline{\theta}\big) = B(z_1, z_2, \dots, z_r, n)\,\frac{\prod_{i=1}^{r}\big[h_i(\underline{\theta})\big]^{z_i}}{\big[g(\underline{\theta})\big]^n} \tag{1.5.32}$$

$z_i \in T(n) \subseteq \mathbb{R}$, $\underline{\theta} \in \Omega$. Here $z = (z_1, z_2, \dots, z_r, n)$ and $B(z_1, z_2, \dots, z_r, n)$ is such that

$$\big[g(\underline{\theta})\big]^n = \int_{z_1 \in T(n)} \int_{z_2 \in T(n)} \cdots \int_{z_r \in T(n)} B(z_1, z_2, \dots, z_r, n) \prod_{i=1}^{r} \big[h_i(\underline{\theta})\big]^{z_i}\, dz_1\, dz_2 \cdots dz_r.$$

Let $\phi\big(\underline{\theta}\big)$ be a function of $\underline{\theta}$ in $f\big(x; \underline{\theta}\big)$ of (1.5.24). Then the following lemma provides a necessary and sufficient condition for the existence of UMVUE of $\phi\big(\underline{\theta}\big)$.

Lemma 1.5.3. *Let $X_1, X_2, \dots, X_n$ be a random sample from (1.5.24). Then there exists the UMVUE of $\phi(\underline{\theta})$ if and only if $\phi(\underline{\theta})[g(\underline{\theta})]^n$ can be expressed in the form*

$$\phi(\underline{\theta})[g(\underline{\theta})]^n = \int_{z_1}\int_{z_2}\cdots\int_{z_r} \alpha(z_1, z_2, \dots, z_r, n) \prod_{i=1}^{r}[h_i(\underline{\theta})]^{z_i}\, dz_1\, dz_2 \cdots dz_r.$$

Proof: *Sufficient* part: Let,

$$\psi(Z_1,\ Z_2, \dots\ , Z_r, n) = \frac{\alpha(Z_1, Z_2, \dots, Z_r, n)}{B(Z_1, Z_2, \dots, Z_r, n)},\ B(Z_1,\ Z_2,\ \dots\ ,\ Z_r, n) \neq 0.$$

Then,

$$E[\psi(Z_1,\ Z_2, \dots\ , Z_r, n)]$$

$$= \int_{z_1}\int_{z_2}\cdots\int_{z_r} \frac{\alpha(z_1, z_2, \dots, z_r, n)\prod_{i=1}^{r}[h_i(\underline{\theta})]^{z_i}}{[g(\underline{\theta})]^n}\, dz_1\, dz_2 \cdots dz_r$$

$$= \phi(\underline{\theta}).$$

Since $\psi(Z_1,\ Z_2, \dots\ , Z_r, n)$ is a function of complete sufficient statistic Z, it is the UMVUE of $\phi(\underline{\theta})$.

Necessary part: Suppose $\psi(Z_1,\ Z_2, \dots\ , Z_r, n)$ is an unbiased estimator of $\phi(\underline{\theta})$. Then,

$$\int_{z_1}\int_{z_2}\cdots\int_{z_r} \frac{\psi(z_1, z_2, \dots, z_r, n) B(z_1, z_2, \dots, z_r, n)\prod_{i=1}^{r}[h_i(\underline{\theta})]^{z_i}}{[g(\underline{\theta})]^n}\, dz_1\, dz_2 \cdots dz_r$$

$$= \phi(\underline{\theta}).$$

That is,

$$\int_{z_1}\int_{z_2}\cdots\int_{z_r}\alpha(z_1,z_2,\ldots,z_r,n)\prod_{i=1}^{r}\left[h_i(\underline{\theta})\right]^{z_i}dz_1\,dz_2\cdots dz_r=\phi(\underline{\theta})\left[g(\underline{\theta})\right]^n$$

where

$$\alpha(z_1,z_2,\ldots,z_r,n)=\psi(z_1,z_2,\ldots,z_r,n)B(z_1,z_2,\ldots,z_r,n).$$

This completes the proof. ■

Thus the UMVUE of a function of $\phi(\underline{\theta})$ of $\underline{\theta}$ is given by

$$\psi(Z_1,\,Z_2,\ldots\,,Z_r,n)=\frac{\alpha(Z_1,Z_2,\ldots,Z_r,n)}{B(Z_1,Z_2,\ldots,Z_r,n)},\ B(Z_1,\,Z_2,\,\ldots\,,\,Z_r,n)\neq 0. \quad (1.5.33)$$

Using the above Lemma 1.5.3, we obtain the UMVUE of different functions of parametric vector $\underline{\theta}$ in the following results:

Result 1.5.8. *Suppose $X_1,X_2,\ldots,X_n$ is a random sample from exponential family defined by (1.5.24). Then,*

$$H_{k_1,k_2,\ldots,k_r}(z_1,z_2,\ldots,z_r,n)=\frac{B(z_1-k_1,z_2-k_2,\ldots,z_r-k_r,n)}{B(z_1,z_2,\ldots,z_r,n)} \quad (1.5.34)$$

is UMVUE of $\prod_{i=1}^{r}\left[h_i(\underline{\theta})\right]^{k_i}$.

Result 1.5.9. *The UMVUE of the variance of $H_{k_1,k_2,\ldots,k_r}(Z_1,\,Z_2,\ldots\,,\,Z_r,n)$ is given by*

$$\widehat{var}\left[H_{k_1,k_2,\ldots,k_r}(z_1,z_2,\ldots,z_r,n)\right]$$
$$=H^2{}_{k_1,k_2,\ldots,k_r}(z_1,z_2,\ldots,z_r,n)-H_{2k_1,2k_2,\ldots,2k_r}(z_1,z_2,\ldots,z_r,n) \quad (1.5.35)$$

Result 1.5.10. *Suppose* $X_1, X_2, \dots, X_n$ *is a random sample from exponential family defined by (1.5.24). Then,*

$$G_k(z_1, z_2, \dots, z_r, n) = \frac{B(z_1, z_2, \dots, z_r, n+k)}{B(z_1, z_2, \dots, z_r, n)} \tag{1.5.36}$$

is the UMVUE of $\left[g\left(\underline{\theta}\right)\right]^k$.

Result 1.5.11. *The UMVUE of the variance of* $G_k(Z_1, Z_2, \dots, Z_r, n)$ *is given by*

$$\widehat{var}[G_k(z_1, z_2, \dots, z_r, n)] = G_k^2(z_1, z_2, \dots, z_r, n) - G_{2k}(z_1, z_2, \dots, z_r, n). \tag{1.5.37}$$

Result 1.5.12. Suppose $X_1, X_2, \dots, X_n$ is a random sample from exponential family defined by (1.5.24). Then,

$$\phi_x(z_1, z_2, \dots, z_r, n) = a(x)\frac{B(z_1 - C_1(x),\, z_2 - C_2(x), \dots,\, z_r - C_r(x), n-1)}{B(z_1, z_2, \dots, z_r, n)} \tag{1.5.38}$$

is the UMVUE of the density $f_X\left(x; \underline{\theta}\right)$.

Result 1.5.13. *The UMVUE of the variance of* $\phi_x(Z_1,\ Z_2,\ \dots\ ,\ Z_r, n)$, $n > 2$ *is given by*

$$\widehat{var}[\phi_x(z_1, z_2, z_3, n)] = \phi_x^2(z_1, z_2, z_3, n)$$
$$-\phi_x(z_1, z_2, z_3, n)\ \phi_x(z_1 - C_1(x),\ z_2 - C_2(x),\ z_3 - C_3(x), n-1) \tag{1.5.39}$$

Corollary 1.5.2. For fixed z, the UMVUE of the survival function $S(t) = P(X > t), t \geq 0$ is given by

$$\hat{S}(t) = \int_{x>t} \phi_x(z_1, z_2, \dots, z_r, n)\, dx. \tag{1.5.40}$$

Result 1.5.14. *For fixed z, the UMVUE of the* $var\left[\hat{S}(t)\right]$ *is given by*

$$\widehat{var}\left[\hat{S}(t)\right] = \hat{S}^2(t) - 2\int_{x>t}\int_{y>t} a(x)a(y)\frac{B(z_1 - C_1(x) - C_1(y), \dots, z_r - C_r(x) - C_r(y)\; n-2)}{B(z_1, \dots, z_r, n)}\,dx\,dy \qquad (1.5.41)$$

.

1.6.Testing of hypothesis for parameters

Since p is the proportion of mixing, it is important to test its presence in the model or not. The hypothesis in this case is to test $H_0\!: p = 1$ against $H_1\!: p < 1$. If the hypothesis is accepted, then it is concluded that there is no mixing and all the observations have come from a failure time distribution stated clearly. In case the hypothesis is rejected, then it is concluded that there is presence of inliers in the model. The detection of number of inliers becomes warranted and this becomes the next stage of inference of such models. The other parameters in the models will be either scale or shape parameters depending upon the FTD, whose presence also are important to assess the overall goodness of the data underlying the mechanism it follows. With this objective, Muralidharan (1999, 2014) has proposed various tests for the model parameters. We briefly discuss a couple of those tests below, and then take up the issue of testing number of inliers.

Consider the inlier prone model $g(x; p, \theta)$, is given in (1.3.7). Using Neyman-Pearson lemma, the most powerful (MP) test for testing $H_0\!: p = 1$ against $H_1\!: p < 1$ of size α is obtained as

$$\Phi_1(r) = \begin{cases} 1, & r < n \\ \alpha, & r = n \\ 0, & r > n \end{cases} \qquad (1.6.1)$$

The above test also uniformly most powerful similar test of size α' with power function $\beta(p) = 1 - (1-\alpha)p^n$, and $\beta(p)$ can be computed numerically for any combination of n, p, and α. And, the locally most powerful (LMP) test of size α for testing $H_0{:}\, p = 1$ against $H_1{:}\, p < 1$ for θ known based on n iid observations from the density $g(x;p,\theta)$ is given by

$$\Phi_2(x) = \begin{cases} 1, & \frac{\partial \log L}{\partial p} < c_\alpha \\ \gamma, & \frac{\partial \log L}{\partial p} = c_\alpha \\ 0, & \frac{\partial \log L}{\partial p} > c_\alpha \end{cases} \tag{1.6.2}$$

where c_α and γ are such that the test attains the level of the test when H_0 is true. i.e. c_α and γ are such that $E_{H_0}[\Phi_2(x)] = \alpha$.

The MP test for $H_0{:}\, \theta = \theta_0$ against $H_1{:}\, \theta = \theta_1$, $(\theta_1 > \theta_0)$ for p known is given by

$$\Phi_3(x) = \begin{cases} 1, & \frac{g_1(x)}{g_0(x)} > c_\alpha \\ \gamma, & \frac{g_1(x)}{g_0(x)} = c_\alpha \\ 0, & \frac{g_1(x)}{g_0(x)} < c_\alpha \end{cases} \tag{1.6.3}$$

where c_α and γ are such that the test attains the level of the test when H_0 is true. i.e. c_α and γ are such that $E_{H_0}[\Phi_3(x)] = \alpha$. The LMP test of size α for testing $H_0{:}\, \theta \le \theta_0$ against $H_1{:}\, \theta > \theta_0$ for p known based on n iid observations from the density $g(x;p,\theta)$ is given by

$$\Phi_4(x) = \begin{cases} 1, & \frac{\partial \log L}{\partial \theta} > c_\alpha \\ \gamma, & \frac{\partial \log L}{\partial \theta} = c_\alpha \\ 0, & \frac{\partial \log L}{\partial \theta} < c_\alpha \end{cases} \tag{1.6.4}$$

where c_α and γ are such that the test attains the level of the test when H_0 is true. That is, c_α and γ are such that $E_{H_0}[\Phi_4(x)] = \alpha$. One may refer to Muralidharan (1999, 2014) for other tests and their comparisons. Below, we discuss the detection of number of inliers in the model.

1.7. Test for number of inliers

1.7.1. Block test procedure for detecting inliers

In literature, two types of tests have been proposed for the lower outlier testing problem. One is the "Block test" and the other one is the sequential procedures. In the "Block test" Chikkagoudar and Kunchur (1983), Kimber and Stevens (1981), and Lewis and Feller (1979), have often used to test the discordancy of k lower outliers in the data, and then declared either k or 0 lower outliers are lower outliers in a single hypothesis test. This procedure suffers from masking and swamping effect when too many or too fewer k inliers are present in the sample. We consider two different inliers prone models to detect inliers in Block test. One is Identified inlier (M_k) model and another is Labeled slippage inlier (L_k) model. They are described below.

***Likelihood ratio test under M_k model*:** Suppose that n units put on a test, let n_0 units fails instantaneously and $(n-n_0)$ failure time is available. Out of these positive observations we have to determine which are inliers or early failures. Before the start of the experiment we do not know which units will fail instantaneously or will produce inliers. These experimental conditions are to be modeled in M_k inliers model for given k. Let us re-label failure times of these $(n-n_0)$ units as $(X_1, X_2, \dots, X_{n-n_0})$. Then in M_k inliers model, we assume that $(n-n_0-k)$ are from target population with pdf $F \in \mathcal{F}$ and k are from the inliers population $G \in \mathcal{G}$. Thus the joint pdf of $(X_1, X_2, \dots, X_{n-n_0})$ can be written as

$$\left(x_1, x_2, \dots, x_{n-n_0} | f, g, v\right) = \prod_{i \in v} g(x_i) \prod_{i \notin v} f(x_i, \theta),\ F \in \mathcal{F}, G \in \mathcal{G}, v \in \nu. \tag{1.7.1}$$

where v is the new parameter representing a set of inliers and ranges over V, the set of integers $(i_1, i_2, \dots, i_k)$ chosen out of $(1, 2, \dots, (n - n_0\ ))$ with cardinality $\binom{n-n_0}{k}$. This is so far similar to the model M_k for k outliers. The main difference in M_k inlier model is that we assume that $G \in \mathcal{G}$ and $F \in \mathcal{F}$ are such that $\frac{\partial G}{\partial F}$ is strictly decreasing in X. The following theorem will help us to write the likelihood function under M_k model.

Theorem 1.7.1. *Let* $X_{(1)} < X_{(2)}, < \cdots < X_{(n-n_0)}$ *be the order statistics and* $(R_1, R_2, \dots, R_{n-n_0})$ *be the corresponding rank order statistics, then* $\underset{1<r_i<\cdots<r_k<n-n_0}{Max} \varphi(r_1, r_2, \dots, r_k) = \varphi(1,2, \dots, k)$ *and* $(x_{(1)}, x_{(2)}, \dots, x_{(k)})$ *have the maximum probability of being inliers.* (See Muralidharan, 2015 page 12.16 for proof).

As a consequence of this theorem, $\hat{\mathrm{v}} = \mathrm{k}$ and therefore, $\mathrm{M_k}\left(\underline{\mathrm{x}} | \mathrm{f}, \mathrm{g}, \hat{\mathrm{v}}\right)$, (say), the likelihood under $\mathrm{M_k}$ inlier model, is

$$M_k\left(\underline{x} | f, g, \hat{v}\right) = A\, k!\, (n - n_0 - k)! \prod_{i=1}^{k} g(x_{(i)}) \prod_{i=k+1}^{n-n_0} f\left(x_{(i)}\right),\ \ F \in \mathcal{F}, G \in \mathcal{G}. \tag{1.7.2}$$

where,

$$A = \binom{n}{n_0} (1 - p)^{n_0} p^{n-n_0}, \; p = P(X > 0). \tag{1.7.3}$$

Suppose we want to test the hypothesis

$H_0 \colon X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are from $F \in \mathcal{F}$

against

$H_1 \colon X_{(1)}, X_{(2)}, \dots, X_{(k)}$ *are from* $G \in \mathcal{G}$ *and*

$X_{(k+1)}, X_{(k+2)}, \dots, X_{(n)}$ *are from* $F \in \mathcal{F}$ (1.7.4)

Then the test is equivalent to testing $H_0: K = 0$ versus $H_1: K = k$. The likelihood ratio using M_k model can be obtained as

$$\Lambda_k(\underline{x}) = \frac{Max\ M_{k0}\ (\underline{x}|f)}{Max M_{k1}\ (\underline{x}|f,g,\hat{\upsilon})} \tag{1.7.5}$$

where $M_{k0}\ (\underline{x}|f)$ and $M_{k1}\ (\underline{x}|f, g, \hat{\upsilon})$ denote the likelihood function under H_0 and H_1 respectiely.

The test reject H_0 if $\Lambda_k(\underline{x}) < C_k(n - n_0, \alpha)$ where $C_k(n - n_0, \alpha)$ is such that

$$P_{H_0}\left(\Lambda_k(\underline{x}) < C_k(n - n_0, \alpha)\right) = \alpha. \tag{1.7.6}$$

***Likelihood ratio test under L_k model*:** In M_k model, $\mathrm{M}_k(\underline{x}|f, g, \hat{\upsilon})$ given by (1.7.2) is likelihood and not the joint pdf of $(x_{(1)}, x_{(2)}, \dots, x_{(n-n_0)})$. Making it a pdf, then, the model for L_k inliers is therefore

$$L_k(\underline{x}|f, g) = A\frac{k!\,(n-n_0-k)!}{\varphi(1,2,\dots,k)}\prod_{i=1}^{k} g(x_{(i)})\prod_{i=k+1}^{n-n_0} f(x_{(i)}), \quad F \in \mathcal{F}, G \in \mathcal{G} \tag{1.7.7}$$

where A is as given in (1.7.3) and $\varphi(1,2, \dots, k)$ is the normalizing constant to make L_k a pdf. The model L_k is called Labeled slippage model and it can also be derived from M_k with $(Y_1, Y_2, \dots, Y_k)$ are iid distributed as $\mathcal{G}$ and $(V_1, V_2, \dots, V_{n-n_{0-k}})$ as iid $\mathcal{F}$ and with the additional condition$(Y_1, Y_2, \dots, Y_k) \leq Min(V_1, V_2, \dots, V_{n-n_{0-k}})$. Therefore, $\varphi(1,2, \dots, k)$ can be interpreted as the probability that the k smallest observations correspond to the order statistics of k inlier observations coming from G and the remaining observations coming from F under the exchangeable model.

$$\varphi(1,2, \dots, k) = \varphi_k(G, F)$$

$$= P\left(x_{(k)} < x_{(k+1)} | G, F\right)$$

$$= \int_{-\infty}^{+\infty} [G(u)]^k \, (n - n_0 - k)[1 - F(u)]^{n-n_0-k-1} \, d\, F(u).$$

(1.7.8)

In order to test the hypothesis shown in (1.7.4), that is $H_0: K = 0$ versus $H_1: K = k$. Then the likelihood ratio using L_k model is obtained as

$$\Lambda_k'\left(\underline{x}\right) = \frac{Max\ L_{k0}\ \left(\underline{x}|f\right)}{MaxL_{k1}\ \left(\underline{x}|f,g,\hat{v}\right)} \tag{1.7.9}$$

where $L_{k0}\ \left(\underline{x}|f\right)$ and $L_{k1}\ \left(\underline{x}|f, g, \hat{v}\right)$ denote the likelihood function under H_0 and H_1 respectively.

The test reject H_0 if $\Lambda_k'\left(\underline{x}\right) < C_k'(n - n_0, \alpha)$ where $C_k'(n - n_0, \alpha)$ is such that

$$P_{H_0}\left(\Lambda_k'\left(\underline{x}\right) < C_k'(n - n_0, \alpha)\right) = \alpha. \tag{1.7.10}$$

1.7.2. Sequential test procedure for detecting inliers

The sequential procedure can be further classified into *inward* and *outward* procedures.

***The inward sequential procedure*:** In the inward sequential procedure, one start with the full sample and single-inlier test applied repeatedly, by starting with most-smallest observation, deleting discordant value at each stage and then applying the test again to the reduced sample. The process continued until a non-significant result is obtained. The estimated number of inliers is $\hat{k}$, is the number of rejected (marginal) tests. In addition to certain theoretical weakness, this inward sequential procedure is not recommended by Kimber (1982) and Chikkagoudar and Kunchur (1983) due to the fact that the presence of two or more lower outliers may well lead to a non-significant result at the very first stage. Due to the limitations in the Block test and the inward sequential procedure suffering from swamping and/or masking effects, Rosner (1975) suggested applying an outward sequential procedure also called "inside-out" sequential procedure to the reduced sample.

***The outward sequential procedure*:** Here, one specifies a maximum number of inliers k, then the k^{th} smallest inlier is being tested first; if this gives a significant result, then k inliers are declared to be discordant; if a non-significant result is obtained, then test the $(k-1)^{th}$ smallest inliers and so on. This process is continued until either a significant result obtained or no inliers can be declared as discordant. This procedure minimizes the probability and magnitude of both masking and swamping effects. As such, the outward procedure has been claimed superior over the inward procedure (see Kimber, 1982, Chikkagoudar and Kunchur, 1983, Balasooriya and Gadag, 1994). However, control of Type 1 error (the probability of a false alarm) is difficult in the outward procedure. Larger the value of k, more the power it loses. The test considers the null hypothesis H_0 that there are no inliers, with multiple alternatives, $H_j, j = 1,2,\dots,k$ that there are j inliers, with test statistic T_j. A single rejection of the k tests rejects the null hypothesis H_0. Thus, to achieve an overall Type I error level of $0 \le \alpha \le 1$, the marginal tests need to have a lower level. Thus, one defines all marginal tests to have equal level b, That is,

$$P_k(T_j > t_j) = b, j = 1,2,\dots,k \tag{1.7.11}$$

and the level b such that

$$P_k(T_j \le t_j, j = 1,2,\dots,k|H_0) = 1 - \alpha. \tag{1.7.12}$$

Clearly, $a^k \leq b \leq a$, where the lower bound corresponds to the case of independent tests (the Bonferonni bound), and the upper bound to perfect dependence. In contrast, for the inward method, the Type I error is equal to the marginal level $(a = b)$. This is because a rejection of the null hypothesis only happens when the first marginal test is rejected. This is the major advantage over the outward procedure in terms of computation and also because no power is lost due to a multiple testing correction.

1.8. Summary of the thesis

The presentation of the thesis is as follows: **Chapter 1** is a detailed introduction of the study and its foundation. Various real-life examples of inliers are discussed in this chapter along with its modeling. An exhaustive literature survey on the study of inliers is presented. It also contains necessary prerequisites for other chapters. We discuss various estimation procedures including the problem of UMVU estimation for inlier prone models. Also, the distributional properties of the complete sufficient statistics are explored in the multi-parameter exponential family. A brief discussion about the test procedures for parameters and number of inliers is also presented.

In **Chapter 2,** we revisit various inliers prone models given in equations (1.3.2), (1.3.3) and (1.3.4) and propose inferential procedures to deal with different life testing situations. We studied the likelihood estimator and its characteristics, also proposed UMVUE for various parametric functions, keeping FTD as an exponential distribution. Various estimators and characteristics are studied using simulation and with the support of numerical examples. The specific examples include one on failure times (in weeks) of 50 items as considered by Murthy et al. (2004) and another one on Vannman (1991) data on drying of woods under schedule 1 of experiment 3.

Chapter 3 reviews the various types of parametric tests for parameters of inliers prone models. We propose the most powerful (MP) test and locally most powerful (LMP) test for parameters p and θ of exponential inliers model. Various test procedures for testing hypothesis consists of single and multiple inliers are reviewed. Some existing tests are revisited and studied in detail. A discussion of data descriptions with inliers proness and comparative study of sequential procedure and block tests is also included in this Chapter. We present the masking effect in Dixon type test and Cochran type test for inliers. The performance of the test is studied based on powers and masking effects. We carried out an extensive Monte Carlo study to investigate the powers and the error probabilities for the effects of masking and swamping effect in the outward test when the number of inliers is more than one. We illustrated the same for two real data on the outward amount of NEFT and rainfall measurement in this chapter.

Chapter 4 we study the inferences of the model given in (1.3.1) by considering the FTD as Lindley distribution. We provide the likelihood estimator of the parameters of model and UMVUE for various parametric functions, including pdf and reliability function along with the SE of estimates. Various estimators and characteristics are studied with two real datasets, one is the NFHS-3 survey on child's age on death for Gujarat state (http://www.dhsprogram.com) and another one is concentration of SO_2 in air for industrial area of Amritsar in month April 2017. Also, we propose the most powerful (MP) test and locally most powerful (LMP) test for parameters p and θ of Lindley inlier model.

In **Chapter 5,** using Type II censored data we carried out the inferential studies of the model (1.3.5) keeping Gompertz distribution as FTD. Apart from MLE, UMVUE, we also studied least squares, weighted least squares, and percentile based estimation procedures for estimating the parameters. We propose the most powerful (MP) test and locally most powerful (LMP) test for parameters p and θ of Gompertz inliers model. The Characteristics of various estimators are studied with the simulation study. An application of inliers prone models is illustrated with a real dataset of Vannman (1991) data on drying of woods under schedule 1 of experiment 2.

The estimation of parameters based on Type II censored sample for model presented in equation (1.3.7) with FTD Weibull distribution is experiment in **Chapter 6**. The Maximum Likelihood Estimators (MLE) are developed for estimating the unknown parameters. The Fisher information matrix, as well as the asymptotic variance-covariance matrix of the MLEs, is derived. UMVUE of model parameters as well as UMVUE of various parametric functions is obtained. The model is implemented on a real data of tumor size in invasive ductal breast carcinoma (IDC) of female patients (www.cbioportal.org). The particular case of the exponential distribution is also included in the chapter as it has a lot of practical significance. The proposed model is then applied on real data based on the NFHS-3 survey for Gujarat state (http://www.dhsprogram.com).

Chapter 7 is devoted to the Type II censored sample for model presented in (1.3.7) with FTD as the Pareto distribution. The Fisher information matrix, as well as the asymptotic variance-covariance matrix of the MLEs, is derived. The UMVUE of model parameters as well as UMVUE of various parametric functions is obtained. The model is implemented on four real datasets of loss ratios for earthquake insurance in California, NFHS-3 data, forest fire burnt are of India 2018, and last three decades average snowfall data in the US. and compared with the Weibull inlier model.

We also provide some discussion on the importance of inliers proness in practical significance in **Chapter 8.** The limitations and future works are also suggested in this chapter.

At the end, we provide the appendix of all datasets and its details used in the thesis and the list of references.