A thesis entitled

Estimation in presence of inliers

By

ARTI M KHABIA

Department of Statistics, Faculty of Science
The M. S. University of Baroda, Baroda.

Submitted to

The Maharaja Sayajirao University of Baroda

Under the guidance of

Prof. K. Muralidharan

Department of Statistics, MSU Baroda.

For the award of the degree of

DOCTOR OF PHILOSOPHY

September, 2011

Certificate

Certified that the work incorporated in the thesis entitled "Estimation in presence of inliers" submitted by Mrs. Arti M. Khabia compromises the results of independent and original investigation carried out by the candidate under my supervision. The material that has been obtained and used from other sources have been duly acknowledged in the thesis.

(Research Guide)

Prof. K. Muralidharan

Professor and Head, Department of Statistics,

M.S. University of Baroda, Vadodara- 390 002.

To

My Son

Kritgya

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my guide Professor K. Muralidharan, the Head, Department of Statistics, The Maharaja Sayajirao University of Baroda, for his insight, guidance, constant encouragement in every manner possible. During my tenure as a doctoral student, I have learned so much from him that words would not be sufficient to express.

I am so indebted to my family, and my parents for their constant support and encouragement throughout my research. Last, but not least, I am deeply grateful to my husband Manish and my son Kritgya, without whose understanding, love and patience, I would not have been able to accomplish my research.

I'll never be able to forget the support and cooperation I received from the Principal of my college, Dr. Basudeb Bakshi. At the same time, I am also very grateful for the assistance and valuable advice I received from all my colleagues at Natubhai V. Patel College of Pure and Applied Sciences, V. V. Nagar, my friends and people who have directly or indirectly helped me in accomplishment of my research.

Date: Arti Khabia

Contents

Chapters

Page No.

Papers Published/ accepted /communicated from this Thesis

1. Analysis of instantaneous and early failures in Pareto distribution, *Journal of statistical theory and Applications,* Vol 7, pg. 187-204, 2008.

2. A modified Pareto distrtibution, *Journal of the Indian Statistical Association,* Vol. 49, No.1, pg. 73-89, 2011.

3. Inliers detection in Pareto distribution, *Journal of Interdisciplinary Mathematics,* Vol. 15, No. 4,5, pg. 261-274, 2012.

4. Detection of inliers in Weibull model, to appear in *Journal of Kerala Association* in 2012.

5. Inlier proness in normal distribution to appear in *International Journal of Systems Assurance Engineering and Management* in 2013.

Chapter 1

Introduction

1.1 Introduction

An inlier in a set of data is an observation or subset of observations not necessarily all zeros, which appears to be inconsistent with the remaining data set. For example: consider the following example as a natural occurrence of a physical phenomenon: 0, 0, 0, 0, 0.01, 0.05, 0.06, 0.71, 1.91, 1.2, 1.76, 2.54, 2.72, 3.07, 3.91 and 3.99. Here the first four observations are instantaneous failures, next three observations may be treated as early failures (by specifying delta δ=0.06 or 0.08) and others may be treated as coming from any positive distribution F. The observations which are identified as instantaneous and early failures together are called inliers, introduced first time by Muralidharan and Kale (2002). In outlier's concept, they may be termed as spurious observations, but unlike outlier concept, we don't discard such observations from analysis and inferences. An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs. In many cases outliers exist in the form of errors of observation or mis-recording due to human errors. Outliers are the surprisingly extreme values occurring on both sides of the distribution whereas inliers occur on left hand side of the distribution. Inliers are integral part of the data and cannot be neglected. For an exhaustive survey and theory of outliers one may refer to Barnett and Lewis (1994) and Rousseeuw and Leroy (1987) the references contained therein. Outlier detection methods have been suggested for numerous applications, such as credit card fraud detection, clinical trials, voting irregularity analysis, data cleansing, network intrusion, severe weather prediction, geographic information systems, athlete performance analysis, and other data-mining tasks. Most of the earliest univariate methods for outlier detection rely on the assumption of an underlying known distribution of the data, which is assumed to be identically and independently distributed (i.i.d.). Moreover, many discordance tests for detecting univariate outliers further assume that the distribution parameters and the type of expected outliers are also known (Barnett and Lewis, 1994). In real world for data-mining applications these assumptions are often violated. In some of the examples discussed above, the inlier observation also becomes a part of outlier observations.

In literature some authors have defined inliers as those observations which are not outliers (Barnett and Lewis, 1994). One can refer Akkaya and Tiku (2005) for this.

Some specific real life situations, where inlier observations are natural occurrences can be described by the following examples:

- In auditing some population elements contain no errors, whereas other population elements contain errors of varying amounts. The distribution of errors can, therefore, be viewed as a mixture of two distinguishable distributions, one with a discrete probability mass at zero and the other a continuous distribution of non-zero positive and/or negative error amounts. The main statistical objective in this auditing problem is to provide a statistical bound for the total error amount in the population.

- In the mass production of technological components of hardware, intended to function over a period of time, some components may fail on installation and therefore have zero life lengths. A component that does not fail on installation will have a life length that is a positive random variable whose distribution may take different forms. Thus, the overall distribution of lifetimes, which includes the duds, is a nonstandard mixture.

- 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 the other a continuous distribution.

- In studies of genetic birth defects, children can be characterized by two variates, a discrete or categorical variable to indicate if one is not affected, affected and born dead, or affected and born alive and a continuous variable measuring the survival time of affected children born alive. The conditional distribution of survival time given, this first variable is undefined for children who are not affected and born dead, and nontrivial for children who are born alive. In some cases it may be necessary to consider the conditional survival time distribution for affected children as a mixture of a mass point at 0 and a nontrivial distribution.

- In measurements of physical performances 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.

- 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, 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.

- Time until remission is of interest in studies of drug effectiveness for treatment of certain diseases. Some patients respond and some do not. The distribution is a mixture of a mass point at 0 and a nontrivial continuous distribution of positive remission times. The problem can be considered for instantaneous failure.

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

- The data recorded for a rainy season can be seen as a combination of zeros (no rainfall) and positive observations (days having nominal or marginal rain reported) etc.

From the above examples, it is seen that the values including zeroes and close to zeroes are important as well as significant in most of the cases. Thus inliers are more natural than the outliers, where most of the time after the detection of outlier(s), the observation(s) may not be considered for further analysis. As a consequence, the modeling of inliers distribution is more important than the detection. Below we discuss some possible models treated in this thesis for detection, estimation and testing.

1.2 Models

Various inlier prone models and their statistical significances are studied in this thesis`. We have considered the following models in various chapters. They are used for analysis of mixture distribution of inliers and target observations, and for estimating the parameters of mixture distribution. Comparison of models are also been done to know which model fits well to the data.

1.2.1 Instantaneous failure model

Consider a model $\Im = \{F(x,\theta), x \geq 0, \theta > 0\}$ where $F(x,\theta)$ is a continuous failure time distribution function (df) with $F(0)=0$. To accommodate a real life situation, where instantaneous failures are observed at the origin, the model $\Im$ is modified to $\mathcal{G} = \{G(x,\theta,\alpha), x \geq 0, \theta \in \Omega, 0 < \alpha < 1\}$ by using a mixture in the proportion $1-\alpha$ and α respectively of a singular random variable Z at zero and with a random variable X with distribution function $F \in \Im$. Thus the modified failure time distribution has the pdf

$$g(x,\theta,\alpha) = \begin{cases} 1-\alpha, & x = 0 \\ \alpha f(x,\theta), & x > 0 \end{cases} \tag{1.2.1}$$

This model has been studied by many authors. The problem of inference about (α,θ) has received considerable attention particularly when X is exponential with mean θ. Some of the early references are Aitchison (1955), Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991), Muralidharan (1999, 2000), Kale and Muralidharan (2000) and references contained therein.

Aitchson (1955) stated the problem of determining efficient estimates of the mean and variance of a distribution specified by (i) non zero probability that the variables assumes a zero value and (ii) a conditional distribution for the positive values of the variable. The estimation problem was analyzed and its implications for the Pearson type-III, exponential, lognormal and Poisson series conditional distribution were investigated. Kleyle and Dahiya (1975) have considered estimation of parameters of mixture distribution of binomial and exponential population. The exact bias and mean square error (MSE) of the estimator is derived and computed for different values of parameters. They had also shown the exact MSE approaches to asymptotic MSE as n increases. Jayade and Prasad (1990) studied the problem of estimation of parameters of a mixture of degenerate and exponential distribution. A new sampling scheme was proposed and the exact bias and MSE of the MLE of the parameters was derived. Moment estimators and their approximate bias and MSE were also obtained. Muralidharan (1999), (2000) obtained tests for the mixing proportion in the mixture of a degenerate rate and exponential distribution. The UMVUE and Bayes estimator of the reliability for some selective prior when the mixing proportion is known and unknown are derived. Muralidharan and Kale (2002) considered the case where *F* is a two parameter Gamma distribution with shape parameter β and scale parameter θ and obtained confidence interval for $\phi=\alpha\beta\theta$ assuming α known and unknown respectively. Singh (2008) obtained UMVUE for mixture of instantaneous and positive observation from exponential families.

1.2.2 Early failure model

To accommodate early failures, the family $\Im$ is modified to new distribution $\mathcal{G}_1 = \{G_1(x,\theta,\alpha), x \geq 0, \theta \in \Omega, 0<\alpha<1\}$ where the d.f. corresponding to $g_1 \in \mathcal{G}_1$ is given by

$$G_1(x,\theta,\alpha) = (1-\alpha)H(x) + \alpha F(x,\theta)$$

where *H*(*x*) is a d.f. with $H(\delta)=1$ for δ sufficiently small and assumed known and specified in advance. The corresponding pdf is then given by

$$g_1(x,\alpha,\theta) = \begin{cases} 0, & x<\delta \\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta \\ \alpha f(x,\theta), & x>\delta \end{cases} \tag{1.2.2}$$

Some of the references which treat early failure analysis with exponential distributions are Kale and Muralidharan (2000), Kale (2001) and Muralidharan (2002), wherein they have treated early failures as inliers using the sample configurations. Muralidharan (2005) has presented in his paper, estimation of parameters in presence of early failures. Kale and Muralidharan (2007) obtained MLE for parameter θ of the target distribution *F* and parameter ϕ of the contaminating population $\mathcal{G}$ assuming number of inliers is known. Muralidharan and Lathika (2008) studied analysis of instantaneous and early failures in Weibull distribution.

Kale and Muralidharan (2008) studied inlier detection using Schwarz information criterion. The estimation of mixture density of inliers and target observation can be viewed as special case of mixture distribution.

1.2.3 Nearly instantaneous failure model

As seen in the data set discussed above, if the observations are closed to zeroes, they can be termed as nearly instantaneous failures. Although the model described in (1.2.2) incorporates inliers for a specified value of δ, there are some drawbacks for the model (1.2.2). This is rectified in the following model as a complete mixture of two distributions. Thus, the nearly instantaneous gives the modification, the density function is given by:

$$f(x) = (1-p) f_1(x) + p f_2(x) \tag{1.2.3}$$

where

$$f_1(x) = \delta_d(x - x_0) = \begin{cases} \dfrac{1}{d}, & x_0 \le x \le x_0 + d \\ 0, & otherwise \end{cases}$$

and $f_2(x)$ can be considered as any other lifetime distribution of target population. A mixture distribution involving two-parameter Weibull distributions has been thoroughly studied by Lai, Khoo, Muralidharan and Xie (2007). The importance of the model is that we can obtain the reliability function and hazard function in closed form. The characteristics of the model, such as survival rate, hazard rate and mean residual life, are studied for various distributions in various chapters for particular cases of $f_2(x)$.

1.2.4 M_k inliers Models and L_k inliers Models

Suppose that *n* units are put on test and n_0 units fail instantaneously and (n-n_0) failure time are available. Out of these positive observations we have to determine which are inliers or early failures. Before the start of the experiment we are unaware of which unit fail instantaneously or will produce early failures. These experimental conditions are to be modeled in M_k inlier model for given k. Let us denote failure times of these (n-n_0) unit as $\left(X_1, X_2, \ldots\ldots X_{n-n_0}\right)$. Then in M_k inlier model, (n- n_0 - k) are considered from target population with pdf $f \in \Im$ and k observations are from inlier population $g \in \mathcal{G}$. Thus the joint pdf of $\left(X_1, X_2, \ldots\ldots X_{n-n_0}\right)$ can be written as

$$L\left(x_1, x_2, \ldots x_{n-n_0} \mid f, g, v\right) = \left\{\prod_{i \in v} g\left(x_i\right)\right\}\left\{\prod_{i \notin v} f\left(x_i, \theta\right)\right\}, \; f \in \Im, v \in V \text{ and } g \in \mathcal{G} \quad (1.2.4)$$

where v is the new parameter representing set of inliers and ranges over V, the set of integers $\left(i_1, i_2, \ldots\ldots i_k\right)$ chosen out of $\left(1,2,\ldots.\left[n-n_0\right]\right)$ and therefore with cardinality $\binom{n-n_0}{k}$. This is so far similar to the model M_k for k outlier. The main difference in M_k inlier model is that $\Psi(x) = \frac{\partial G}{\partial F} = \frac{g(x)}{f(x)}$ is assumed to be strictly decreasing function of X. The theorem stated below is used to write the likelihood function under M_k and L_k reproduced from Muralidharan (2010), for continuity.

Theorem 1.2.1: Let $\left(X_{(1)} < X_{(2)} < \ldots\ldots < X_{\left(n-n_0\right)}\right)$ be the order statistics observations and $\left(R_1, R_2, \ldots. R_{n-n_0}\right)$ be the corresponding rank order statistics then $Max\, \varphi\left(r_1, r_2, \ldots. r_k\right) = \varphi\left(1,2,\ldots\ldots k\right)$ and $x_{(1)}, x_{(2)} \ldots\ldots x_{(k)}$ have the maximum probability of being inliers.

Here we give only the important outline of the theorem. Assume that model contains n-n_0 positive observations. Then for one inlier model considered is

Proof: Consider M_1 and $P\left[R_1 = r_1, X_{\left(r_1\right)} = X_{ir_1} \mid x_{ir_1}, g\right] = \varphi\left(r_1\right)$. Then

$$\varphi\left(r_1\right) = \binom{n-n_0-1}{r_1-1} \int \left[F(x)\right]^{r_1-1} \left[1-F(x)\right]^{n-n_0-r_1} dG(x)$$

now

$$\Psi(x)=\frac{\partial G}{\partial F}=\frac{g(x)}{f(x)},$$

therefore,

$$\varphi(r_1)=\binom{n-n_0-1}{r_1-1}\int_0^{\infty} y^{r_1-1}\left[1-y\right]^{n-n_0-r_1}\Psi\left[F^{-1}(y)\right]dy$$

$$=\frac{1}{n-n_0}E\left[\Psi_1(y_r)\right] \qquad (1.2.5)$$

Now y_r is a beta random variable with parameters $(r_1, n-n_0-r_1+1)$. Note that, $r_1=1,2,\ldots n-n_0$ is stochastically ordered sequence, since h is such that $\frac{h(y_r+1)}{h(y_r)}\,\alpha\,\frac{y}{1-y}$ which is strictly increasing function of y over (0,1). $\Psi\left[F^{-1}(y)\right]$ is decreasing function of y by as per our assumption. Therefore, from the result of Lehman (1959) it follows that $\varphi(1)>\varphi(2)>\ldots..>\varphi(n)$ and $X_{(1)}$ has maximum probability of being an inlier. Let $\varphi(r_1,r_2)=$Probability that $X_{(r_1)}$ and $X_{(r_2)}$ are inliers for $1\le r_1\le r_2\le n.$ for model M_2, where

$$\varphi(r_1,r_2)=\frac{(n-n_0-2)!2!}{(r_1-1)!(r_2-r_1-1)!(n-n_0-r_2)!}\int\int_{0<x<y<\infty}\left[F(x)\right]^{r_1-1}\left[F(y)-F(x)\right]^{r_2-r_1-1}\left[1-F(y)\right]^{n-n_0-r_2}dG(x)dG(y)$$

$$=\frac{(n-n_0-2)!2!}{(n-n_0-r_2)!}\int_0^1\left[\int_0^v\frac{u^{r_1-1}(v-u)^{r_2-r_1-1}\Psi\left[F^{-1}(u)\right]}{(r_1-1)!(r_2-r_1-1)!}du\right]\left[1-v\right]^{n-n_0-r_2}\Psi\left[F^{-1}(v)\right]dv$$

Then one can show $Max_{r_1<r_2}\varphi(r_1,r_2)=\varphi(1,2)$ and $x_{(1)}$ and $x_{(2)}$ have maximum probability of being inliers.

Generalizing the above result we can show that

$Max_{r_1\le r_2\le\ldots\ldots\le r_k\le n-n_0}\varphi(r_1,r_2,\ldots.r_k)=\varphi(1,2,\ldots\ldots k)$ and hence $x_{(1)},x_{(2)}\ldots\ldots x_{(k)}$

have the maximum probability of being inliers.

For other detailed proof of the theorem, one may refer to the paper by Muralidharan (2010). Thus the generalized form of $\varphi(\bullet)$ with k inliers is

$$\varphi(r_1,r_2,....r_k)=\frac{(n-n_0-k)!k!}{(r_1-1)!(r_2-r_1-1)!.....(n-n_0-k)}\int_{0<w_1<w_2<....w_k<1}\Big\{w_1^{r_1-1}\left[w_2-w_2\right]^{r_2-r_1-1}......$$

$$....\left[1-w_k\right]^{n-n_0-n_k}\Psi\left[F^{-1}(w_1)\right]......\Psi\left[F^{-1}(w_k)\right]\Big\}dw_1dw_2.....dw_k$$

Now fixing $(r_2,r_3,...,r_k)$ and $(w_2,w_3,...,w_k)$ we can show that $\phi(r_1,r_2,...,r_k)$ as decreasing function of r_1 for $1\le r_1\le r_2$.

Thus the model for M_k inlier is

$$L(x\,|\,g,f,\hat{v})=\prod_{i=1}^{k}g\left(x_{(i)}\right)\prod_{i=k+1}^{n-n_0}f\left(x_{(i)}\right),\ f\in\Im,g\in \mathrm{G}, \tag{1.2.6}$$

But $L(x\,|\,g,f,\hat{v})$ is likelihood and not joint pdf of $x_{(1)},x_{(2)}......x_{(n-n_0)}$.

The model for L_k inlier is therefore

$$L(\underline{x}\,|\,g,f)=\frac{(n-n_0)!k!}{\phi_k(F,G)}\prod_{i=1}^{k}g\left(x_{(i)}\right)\prod_{i=k+1}^{n-n_0}f\left(x_{(i)}\right),\ f\in\Im,g\in \mathcal{G}, \tag{1.2.7}$$

where $\varphi_k(F,G)=\varphi(r_1,r_2,...,r_k)$ is the normalizing constant to make L_k a pdf. The model is called as labeled slippage model and it can be derived as model from M_k with $(Y_1,Y_2,...,Y_k)$ are i.i.d.as $\mathcal{G}$, and $\left(V_1,V_2,...,V_{n-n_0}\right)$ as i.i.d from $\Im$ and with the additional condition $Max(Y_1,Y_2,...,Y_k)\le Min\left(V_1,V_2,...,V_{n-n_0}\right)$.

1.3 Information criteria for inliers

The most important use of information criterion is, that it helps us in model selection, from the set of different models which all fit the data. These criterion are suitable when the underlying distribution and inlier distribution are available. It is an exploratory data analysis approach as no formal statistical inference is performed. The Akaike information criterion is a measure of the relative goodness of fit_of a statistical model. It was developed by Hirotsugu Akaike, under the name of "an information criterion" (*AIC*), and was first published by Akaike (1974). It is grounded in the concept of information entropy, in effect offering a relative measure of the information lost when a given model is used to describe reality. It can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking between accuracy and complexity of the model. In statistics, the Bayesian information criterion (*BIC*) or Schwarz criterion (also *SBC*, *SBIC*) is a criterion for model selection among a class of parametric models with different numbers of parameters. Choosing a model to optimize *BIC* is a form of regularization. When estimating model parameters using maximum likelihood estimation, it is possible to increase the likelihood by adding parameters, which may result in overfitting. The *BIC* resolves this problem by introducing a penalty term for the number of parameters in the model. This penalty is larger in the BIC than in the related *AIC*. The *BIC* was developed by Gideon E. Schwarz (1978), who gave a Bayesian argument for adopting it. It is closely related to the Akaike information criterion (*AIC*). In fact, Akaike was so impressed with Schwarz's Bayesian formalism that he developed his own Bayesian formalism, now often referred to as the *ABIC* for "a Bayesian Information Criterion" or more casually "Akaike's Bayesian Information Criterion". The BIC is an asymptotic result derived under the assumptions that the data distribution is in the exponential family.

Given any two estimated models, the model with the lower value of *BIC* is the one to be preferred. The *BIC* is an increasing function of σ_e^2 (variance) and an increasing function of *p,* where *p* is number of parameters of population under study. That is, unexplained variation in the dependent variable and the number of explanatory variables increase the value of *BIC*. Hence, lower BIC implies either fewer explanatory variables, better fit, or both. The *BIC* generally penalizes free parameters more strongly than does the Akaike information criterion, though it depends on the size of *n* and relative magnitude of *n* and *k*. It is important to keep in mind that the *BIC* can be used to compare estimated models only when the numerical values of the dependent variable are identical for all estimates being compared. The models being compared need not be nested, unlike the case when models are being compared using an *F* or likelihood ratio test.

The following information criteria are used in all the chapters. The Schwarz's Information criterion as given by $SIC = -2\ln L(\Theta) + p\ln n,$ Schwarz's Bayesian Information criterion as obtained by $BIC = -\ln L(\Theta) + \frac{0.5(p\ln n)}{n}$ and Hannan-Quinn criterion as given by $HQ = -\ln L(\Theta) + p\ln(\ln(n))$ to detect the inliers, where L(Θ) the maximum likelihood function and *p* is the number of free parameters that need to be estimated under the model.

Before discussing the tests of hypothesis we provide another theorem, again reproduced from Kale and Muralidharan (2007), which will help to understand the inlier distribution from among the other distributions. If *F* and *G* are respectively given by

$$F(x,\theta) = 1 - \exp(-x\theta), \quad x > 0, \theta > 0,$$

and

$$G(x,\phi) = 1 - \exp(-x\phi), \quad x > 0, \phi > 0 \text{ where } \phi = \lambda\theta,\ \lambda > 0$$

Using theorem (1.2.1), the labeled slippage alternative of $r \geq 1$ are discordant observation H_r, the joint distribution of the ordered statistics is given by

$$f\left(x_{(1)}, x_{(2)} \ldots . x_{(n)} \mid H_r\right) = \frac{(n-r)!\,r!\,\lambda^r}{\varphi(1,2,\ldots r)} \exp\left\{-\lambda \sum_{i=1}^{r} x_{(i)} - \sum_{i=k+1}^{n} x_{(i)}\right\} \quad (1.3.1)$$

where the normalizing factor is given by

$$\varphi(1,2,\ldots r) = \frac{n-r}{\lambda} B\left[r+1, \frac{n-r}{\lambda}\right], \lambda > 0, r \geq 1.$$

Then we have the following theorem

Theorem 1.3.1: Under the labeled slippage alternative, H_r, $\frac{x_{(1)}}{x_{(i)}} \xrightarrow{P} 0$ as $\lambda \to \infty$, for $i = r+1, r+2, \ldots, n$.

Proof: From (1.3.1) the joint density of $X_{(1)}$ and $X_{(r+1)}$ can be obtained as

$$f\left(x_{(1)}, x_{(r+1)}\right) = \frac{(n-r)r\lambda}{\varphi(1,2,\ldots,r)} e^{-\lambda x_{(1)}} \left[e^{-\lambda x_{(1)}} - e^{-\lambda x_{(r+1)}}\right]^{k-1} e^{-\left(n-r x_{(r+1)}\right)} \tag{1.3.2}$$

where $0 \le x_{(1)} \le x_{(2)}$

$$= \frac{(n-r) r e^{-\left(n-r x_{(r+1)}\right)}}{k\varphi(1,2,\ldots,r)} \left(e^{-\lambda x_{(1)}} - e^{-\lambda x_{(r+1)}}\right)^{k}$$

$$f\left(x_{(1)} \mid x_{(r+1)}\right) = \frac{\lambda e^{-\lambda x_{(1)}}}{k\varphi(1,2,\ldots,r)} \frac{\left(e^{-\lambda x_{(1)}} - e^{-\lambda x_{(r+1)}}\right)^{k-1}}{\left(e^{-\lambda x_{(2)}} - e^{-\lambda x_{(r+1)}}\right)^{k}}$$

hence for all $a \in (0,1)$, we get

$$P\left(\frac{X_{(1)}}{X_{(r+1)}} < a \mid H_r\right) =$$

$$\frac{(n-r)r\lambda}{\varphi(1,2,\ldots,r)} \sum_{i=0}^{r-1} \frac{(-1)^i \binom{r-1}{i}}{\left[\lambda(r-1-i)+(n-1)\right]\left\{\frac{1}{a}\left[\lambda(r-1-i)+(n-r)\right]+\lambda(i+1)\right\}}$$

One gets $\frac{x_{(1)}}{x_{(r+1)}} \to 0$ as $\lambda \to \infty$. which proves the theorem given the condition

$0 \le \frac{x_{(1)}}{x_{(i)}} \le \frac{x_{(1)}}{x_{(r+1)}} \Rightarrow X_{(r+1)} < X_{(i)}$, $i = r+2, \ldots, n$.

1.4 Testing of hypothesis

The main objective of this thesis is to detect (estimate) number of inliers in a given data. After detecting the number of inliers, using some model, we subjected the finding to test whether our results are true in light of a random sample. For which we have used various test to do this. Some Traditionally used tests are discussed below: In most of the test procedure, our main objective is to test the hypothesis:

$H_0 : X_{(1)}, X_{(2)}...X_{(n)}$ are from $F(x, \theta)$ and

$H_1 : X_{(1)}, X_{(2)}...X_{(r)}$ are from $G(x, \phi)$ and $X_{(r+1)}, X_{(r+2)}...X_{(n)}$ are from $F(x, \theta)$, (1.4.1)

For a hypothesis of the form in equation (1.4.1) one can construct likelihood ratio test for testing inliers in the usual way. For example the underlying density is exponential, then the likelihood ratio test for one inlier by Kale and Muralidharan (2007) is obtained as

Reject H_0 if

$$\frac{T}{x_{(1)}} > c \ , \tag{1.4.2}$$

where $T = \sum_{i=1}^{n} X_{(i)}$. And the value of $c = \frac{n}{1-(1-\alpha)^{\frac{1}{n-1}}} - 1.$

Also the power of the test for one inlier is given by

$$P_1(\lambda) = 1 - \left(\frac{c-n+1}{c+\lambda}\right) \quad \text{where} \quad \lambda = \frac{\theta}{\phi} \tag{1.4.3}$$

Specifically if $X_1, X_2, ...X_n$ are independent and identically distributed r.v's having mixture distribution with likelihood is

$$L(x,\phi,\theta,p) = \prod\{(1-p)g(x_i) + pf(x_i)\} \tag{1.4.4}$$

then the objective is to test

$H_0 : p = 1$ against $H_1 : p < 1$

for which we can have the following tests:

1.4.1 Locally most powerful test

The LMP test critical region for equation (1.4.4) is given by

$$\left[\underline{x} \mid \frac{\partial L(x,\phi,\theta,p)}{\partial p} \mid H_0\right] \leq C \tag{1.4.5}$$

where *C* is such that

$$P\left\{\left[\underline{x} \mid \frac{\partial L(x,\phi,\theta,p)}{\partial p} \mid H_0\right] \leq C\right\} = \alpha, \text{ the size of test.}$$

1.4.2 A Large sample test

A large sample test for the hypothesis (1.4.4) can be constructed using the asymptotic binomial distribution of the parameter of *p*: The large sample test for the hypothesis

$H_0 : p \geq p_0$ against $H_1 : p < p_0$, p_0 specified.

The test statistics is given by

$$Z_{cal} = \frac{\sqrt{n}\,(\hat{p} - p_0)}{\sqrt{p_0 q_0}}, \quad q_0 = 1 - p_0 \tag{1.4.6}$$

and we reject H_0 if $Z_{cal} < Z_\alpha$ where α is level of significance. *p* denotes proportion of observations from target population.

1.5 Inlier estimation through Sequential Probability Ratio Test (SPRT)

Here first we want to test the hypothesis whether an observation belongs to inliers population with pdf $g(x,\phi)$ against hypothesis that it belongs to target population with pdf $f(x,\theta)$.

That is if $L_1 = \prod_{i=1}^{r} f\left(x_i, \theta\right)$ and $L_0 = \prod_{i=1}^{r} g\left(x_i, \phi\right)$ denote likelihood function under target and inlier population respectively, then the SPRT is the likelihood ratio λ_r is given by

$$\lambda_r = \frac{L_1}{L_0}$$

or equivalently

$$\ln \lambda_r = \sum_{i=1}^{r} \ln \frac{f\left(x_{(i)}, \theta\right)}{g\left(x_{(i)}, \phi\right)} = \sum_{i=1}^{r} z_{(i)} \qquad r = 1,2,....n \tag{1.5.1}$$

For deciding number of inliers r we continue to take additional observations till we reject H_0. That is

if $\sum_{i=1}^{r} z_{(i)} \leq \ln B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{r} z_{(i)} \geq \ln A$ reject H_0 and stop.

The corresponding r represents the first observation from $f\left(x_{(i)}, \theta\right)$ and the previous (r-1) observations from $g\left(x_{(i)}, \phi\right)$. Thus the number of inliers will be r - 1.

1.6 Most powerful test for detection of inliers when underlying parameters are specified

If we are interested in testing $H_0 : g\left(x_{(i)}, \phi\right)$ against $H_1 : f\left(x_{(i)}, \theta\right)$ (i.e. whether data is from inlier population against data is from target population) a MPT can be constructed for ξ, the common parameter of interest, then the hypothesis can be equivalently written as

$$H_0 : \xi = \phi \text{ against } H_1 : \xi = \theta. \tag{1.6.1}$$

In the above frame, both H_0 and H_1 are simple and hence the most powerful test according to NP lemma is

$$\psi(x)=\begin{cases}1, & \dfrac{P_1(x)}{P_0(x)}>C\alpha \\ 0, & \dfrac{P_1(x)}{P_0(x)}<C\alpha\end{cases} \tag{1.6.2}$$

where the constant $C\alpha$ can be obtained using the size condition. For specific distributional model, the value of $C\alpha$ can be numerically computed.

In chapters to follow, we have studied many other test procedures and interesting properties of the models. For situation specific, we have changed the notation and theoretical development to establish proper continuity. We now provide the chapter wise summary of the thesis, in brief.

Chapter 1 gives a detailed introduction of the study and its need. An exhaustive literature survey on study of inliers is discussed. The utility and applicability of inlier distributions are also discussed in length and breadth. Various real life examples and their application areas are discussed in this chapter.

Chapter 2 discusses Pareto distribution as a inlier model for file sizes on the internet, insurance losses, and financial behavior of the stock market and in telecommunication systems. The proposed study is a further look at suitability of Pareto distributions in the context of life testing experiments where data involves instantaneous and early failures. We provide the inferences on parameters of modified forms of Pareto type distributions involving one and two parameters. The methods are illustrated on simulated data sets and on a real life data. We have discussed different criteria for detection of inliers and studied the sensitivity of various distributions with respect to different hypothesis. Through many other characteristics, we have shown that the Pareto distribution is much better than that of the Weibull distribution, in identifying inliers, and inlier models

In **Chapter 3,** we study the estimation of inliers in Normal distribution. The masking effect problem for correctly identifying the inliers is discussed with respect to various test procedures. Test for detecting a single inlier, H_o against H_1 is based on symmetric functions of observations or on functions of order statistics. In the *k*-inlier model, the joint distribution of order statistics $X_{(1)}, X_{(2)}, \ldots . X_{(n)}$ is same as that under the exchangeable model introduced by Kale (1998) where it is assumed that any set $X_{i_1}, X_{i_2}, \ldots . X_{i_k}$ has priori equal probability of being independent and identically distributed as G_λ and the remaining (*n*-*k*) observation are distributed as *F*, the distribution function of target population.

The study of inliers in Weibull models is the content of **chapter 4**. Apart from the regular estimation of inliers, we have also discussed the model specific estimation when the total realizations are assumed to be from either Model-1 or Model-2. If we assume, the data $\underline{X}=(x_1, x_2, \ldots\ldots x_n)$ whose joint distribution is unknown, and if we have two competing models with parametric density $f_j(x;\theta_j,\alpha)$, $\theta_j \in \Theta_j$, where Θ_j is the Parametric space. Model-1 is selected with inliers and target population both having Weibull distribution with same shape parameter whereas other model-2 has Weibull with same scale parameter. We have also used predictive approach to model selection using exponential model. The SPRT test is conducted to detect number of inliers in both the models. Conditional test and Predictive method are also incorporated to detect inliers in exponential models.

In **chapter 5** we study the usefulness of mixture distribution and modified distribution for inlier study. Mixture distributions have been extensively used in a wide variety of important practical situations where data can be viewed as arising from two or more populations mixed in varying proportions. Mixture of distribution refers to the situation in which i^{th} distribution out of k underlying distribution is chosen with probability p_i ,i=1,2,….*k*. Mixture distribution having *k*=2 components are extensively studied in literature. For example a probability model for the life of a electronic product can be described as the mixture of two unimodel distribution, one representing the life of inliers and other for target observations. We have listed down the methods which will be useful in detecting inliers present in the sample data. The graphs representing mixture of inliers and target populations, for exponential families are also plotted.

The inlier detection in generalized distribution is included in **chapter 6.** A generalized treatment for estimation and detection of inliers is discussed in this chapter. We also studied estimation of parameters of mixture distribution for particular cases. Apart from this we have derived the test for one inlier in the data set.

At the end we have given an exhaustive and extensive bibliography. As an output of the thesis, two articles have been published and couple of papers is on the way for publication. About three papers are ready for communication.

Chapter 2

Inlier estimation in Pareto distribution

2.1 Introduction

Pareto distribution has recently been used as a model for file sizes on the internet, insurance losses, and financial behavior of the stock market as well as in telecommunication systems. Many of the empirical studies also use Pareto's law for representing long tail distributions. The proposed study is aimed to look further for the suitability of Pareto distributions in the context of life testing experiments where data involves instantaneous and early failures. The occurrence of instantaneous or early failures in life testing experiment is a phenomenon observed in electronic components as well as in clinical trials. These occurrences may be due to inferior quality or faulty construction or due to no response of the treatments. Such failures usually discard the assumption of a unimodal distribution and hence the usual method of modeling and inference procedures may not be accurate in practice. These situations can be handled by modifying used parametric model Pareto distribution. The modified model is then a non-standard mixture of distribution by mixing a singular distribution at zero to accommodate instantaneous failures.

Consider a model $\Im=\{F(x;\theta),x\geq 0,\theta\in\Omega\}$ where $F(x,\theta)$ is a continuous failure time distribution function (df) with $\mathrm{F}(0)=0$. To accommodate a real life situation, where instantaneous failures are observed at the origin, the model $\Im$ is modified to $\mathcal{G}$ = $\{G(x;\alpha,\theta)=(1-\alpha)+\alpha f(x,\theta),0<\alpha<1,F\in\Im\}$ by using a mixture in the proportion 1-α and α respectively of a singular random variable Z at zero and with a random variable X with df $F\in\Im$. The df corresponding to G$\in$ $\mathcal{G}$ is given by

$$\mathrm{G}(x,\theta,\alpha) = (1-\alpha)+\alpha\mathrm{F}(x,\theta) \tag{2.1.1}$$

Thus the modified failure time distribution will have its corresponding probability density function (pdf) as

$$g(x;\theta,\alpha)=\begin{cases}1-\alpha, & x=0\\ \alpha f(x,\theta), & x>0\end{cases} \tag{2.1.2}$$

The problem of inference about (α,θ) has received considerable attention particularly when X is exponential with mean θ. Some of the early works are by Aitchison (1955), Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991), Muralidharan (1999, 2000), Kale and Muralidharan (2000) and references contained therein. Vannman (1995) and Muralidharan and Kale (2002) considered the case where F is a two parameter Gamma distribution with shape parameter β and scale parameter θ and obtained confidence interval for $\phi=\alpha\beta\theta$ assuming α known and unknown respectively.

To accommodate early failures, the family $\Im$ is modified to $\mathcal{G}_1$ = $\{G_1(x,\theta,\alpha),\ x\geq 0,\ \theta\in\Omega,\ 0<\alpha<1\}$ where the df corresponding to $G_1\in$ $\mathcal{G}_1$ is given by

$$G_1(x,\theta,\alpha)=(1-\alpha)H(x)+\alpha F(x,\theta) \tag{2.1.3}$$

where $H(x)$ is a df with $H(\delta)$ = 1 for δ sufficiently small and assumed to be known and specified in advance. We also assume that the early failures are recorded as a class with notional failure time δ so that the modified family G_1 has a pdf with reference to measure μ which is sum of Lebesgue measure on (δ, ∞) and a singular measure at δ. The corresponding pdf is then given by

$$g_1(x,\alpha,\theta)=\begin{cases}0, & x<\delta\\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta\\ \alpha f(x,\theta), & x>\delta\end{cases} \tag{2.1.4}$$

Some of the references which treat early failure analysis with exponential distribution are Kale and Muralidharan (2000), Kale (2001), Kale and Muralidharan (2002), and Muralidharan and Lathika (2006), wherein they treat early failures as inliers using the sample configurations.

The objective is to consider the model *G* given by (2.1.1), G_1 given by (2.1.3) and nearly instantaneous failures when *F* is Pareto and study the suitability of Pareto distribution in the context of life testing experiments. The Pareto distribution was originally derived in connection with studying income distribution. The Pareto distribution is a power-tailed distribution which is a special case of a heavy-tailed distribution whose tails go to zero more slowly than exponential. Many of the empirical studies also use Pareto's law for representing long tail distributions. The distribution also comes in various forms and types. Hence modeling differences between one parameter, two parameters and three parameters Pareto will be a point of interest. Fisher, Masi, Gross and Shortle (2005) have studied the modeling difference of such different forms of Pareto distribution in connection with queuing systems. A three parameter Pareto type family has the survival function

$$\bar{F}(x)=\left(\frac{\beta}{x+\beta-\gamma}\right)^{\phi}, x\geq\gamma, \phi>0, \beta>0, \gamma>0 \tag{2.1.5}$$

or the more general form

$$\bar{F}(x)=\frac{\beta^{\phi}}{\beta+(x-\gamma)^{\phi}}, x\geq\gamma, \phi>0, \beta>0, \gamma>0 \tag{2.1.6}$$

A two parameter Pareto can be easily obtained as a particular case of the above distributions for $\gamma = 0$. The other forms of Pareto can be easily obtained for particular cases of β and ϕ.

We study two types of Pareto distribution in the context of instantaneous failures and early failures. From the point of view of estimating equations, Kale and Muralidharan (2000) have shown that $I_g^{(\alpha)}(\theta)$, Fisher information about θ ignoring α in the model G is less than $I_f(\theta)$, the Fisher information about θ in the original model $\Im$. It is also shown that the parameter α is orthogonal to θ in the case of model (2.1.1), whereas, the parameter α is not orthogonal to θ in the case of model (2.1.3). It is possible to show $I_g^{(\alpha)}(\theta) < I_f(\theta)$ although Var (X|g) can be smaller than Var (X|f) in both the models. In the subsequent sections, different types of Pareto distributions have been used with different parameters for analysis. The general theory of estimating equations and Fisher information's for instantaneous failures and early failures have been developed in the next two sections separately. We also discuss the importance of instantaneous and early failures in practical situations through a real life data set obtained by Vannman (1991).

2.2 Analysis for instantaneous failures

In this section we study inference regarding instantaneous failure. We have obtained UMVUE, Fishers information and MLE for the parameters of inlier and target population.

2.2.1 Fisher information

The pdf in (2.1.2) is with respect to the measure $\mu(x)$ which is the sum of Lebesgue measure over $(0,\infty)$ and a singular measure at $\{0\}$. If we assume $f\left(x_{(i)},\theta\right)$ as a Cramer family, then ln $[g(x,\theta,\alpha)]$ admits continuous partial derivatives with respect to $(\alpha,\theta)'$ upto order two. Here θ can also be a vector of parameters. Further,

$\int_0^\infty g(x,\theta,\alpha)d\mu = 1$, can be differentiated twice under integral sign with respect to $(\alpha,\theta)'$. Hence G satisfies all the regularity conditions of Cramer (1966) and G is a Cramer family. Therefore from (2.1.2),

$$\frac{\partial \ln g}{\partial \alpha} = \begin{cases} \dfrac{-1}{(1-\alpha)}, & x=0 \\ \dfrac{1}{\alpha}, & x>0 \end{cases}$$

$$\frac{\partial \ln g}{\partial \theta} = \begin{cases} 0, & x=0 \\ \dfrac{\partial \ln f(x,\theta)}{\partial \theta}, & x>0 \end{cases}$$

One can verify that $E\left(\dfrac{\partial \ln g}{\partial \alpha}\right)=0$ and $E\left(\dfrac{\partial \ln g}{\partial \theta}\right)=0$. The element of the Fisher information matrix, $I_g(\alpha,\theta)$ are

$$I_{\alpha\alpha} = E\left(-\frac{\partial^2 \ln g}{\partial \alpha^2}\right) = \frac{1}{\alpha(1-\alpha)} \tag{2.2.1}$$

$$I_{\theta\theta} = E\left(-\frac{\partial^2 \ln g}{\partial \theta^2}\right) = \alpha I_f(\theta) \tag{2.2.2}$$

and

$$I_{\alpha\theta} = I_{\theta\alpha} = E\left(-\frac{\partial \ln g}{\partial \beta}\frac{\partial \ln g}{\partial \theta}\right) = \int_{S-\{0\}} \frac{\partial \ln f}{\partial \theta}\alpha f(x,\theta)dx = 0 \tag{2.2.3}$$

Hence $I_g(\alpha,\theta)$ = $diag\left(\dfrac{1}{\alpha(1-\alpha)}, \alpha I_f(\theta)\right)$, which shows that α and θ are orthogonal parameters. Using the definition of Fisher information for θ ignoring α in model $g \in \mathcal{G}$ as given by Liang (1983), denoted by $I_g^{(\alpha)}(\theta)$, we have

$$I_g^{(\alpha)}(\theta) = I_{\theta\theta} - I_{\theta\alpha} I_{\alpha\alpha}^{-1} I_{\alpha\theta} \tag{2.2.4}$$

$$= I_{\theta\theta} = \alpha I_f(\theta) \quad \text{as} \quad I_{\alpha\theta} = I_{\theta\alpha} = 0$$

Since $0 < \alpha < 1$, $I_g^{(\alpha)}(\theta) < I_f(\theta)$ and there is less information about θ ignoring α in the model $\mathcal{G}$ than that in the model $\Im$.

2.2.2 Maximum likelihood estimation.

Now let ($X_1, X_2, ..., X_n$) be a random sample of size *n* from g$\in G$ and define

$$z(x) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases}$$

Then the likelihood function can be written as

$$\begin{aligned} L(x;\alpha,\theta) &= \prod_{i=1}^{n} g(x_i,\alpha,\theta) \\ &= \prod_{i=1}^{n} (1-\alpha)^{z(x_i)} [\alpha f(x_i,\theta)]^{1-z(x_i)} \\ &= (1-\alpha)^{\sum z(x_i)} \alpha^{n-\sum z(x_i)} \prod_{x_i>0} f(x_i,\theta) \end{aligned} \quad (2.2.5)$$

If $\sum_{i=1}^{n} z(x_i) = n_0$, then the likelihood equations are given by

$$\frac{\partial \ln L}{\partial \alpha} = \frac{-n_0}{1-\alpha} + \frac{n-n_0}{\alpha} = 0 \quad (2.2.6)$$

and

$$\frac{\partial \ln L}{\partial \theta} = \sum_{x_i>0} \frac{\partial \ln f(x_i,\theta)}{\partial \theta} = 0 \quad (2.2.7)$$

then from (2.2.6), we have $\hat{\alpha} = \frac{n-n_0}{n}$ and $\hat{\theta}$ will be the solution of (2.2.7). Using the standard results on MLE, we have $\frac{1}{n} I_g^{-1}(\alpha,\theta) = diag\left(\frac{\alpha(1-\alpha)}{n}, \frac{1}{n\alpha I_f(\theta)}\right)$.

The pdf of one parameter Pareto is defined as below

$$f(x)=\frac{\beta x^{\beta-1}}{(1+x^{\beta})^2}, x>0, \beta>0$$

$$\int_0^\infty\left[\frac{x^{\beta}(\ln x)^2}{(1+x^{\beta})^2}\right]f(x)dx=\frac{1}{\beta^2}\frac{(\pi^2-6)}{18}$$

where the log likelihood is

$$\ln L=r\ln(1-\alpha)+(n-r)[\ln\alpha+\ln\beta]+(\beta-1)\sum_{x_i>0}\ln x_i-2\sum_{x_i>0}\ln\left(1+x_i^{\beta}\right)$$

and the Fisher information's are

$$I_{\alpha\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha^2}\right)=\frac{1}{\alpha(1-\alpha)}$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^2\ln L}{\partial\alpha\partial\beta}\right)=0$$

and

$$I_{\beta\beta}=E\left(-\frac{\partial^2\ln L}{\partial\beta^2}\right)=\frac{\alpha}{\beta^2}\left[1+\frac{(\pi^2-6)}{9}\right].$$

The pdf for two parameter Pareto is as given below

$$f(x)=\frac{\beta}{\phi}\left(\frac{x}{\phi}\right)^{\beta-1}\frac{1}{\left[1+(x/\phi)^{\beta}\right]^2},\ x\geq 0,\ \ \phi>0,\ \beta>0$$

For some computations below, we use the following formulas:

$$\int_0^\infty\left[\frac{(x/\phi)^{\beta}}{\{1+(x/\phi)^{\beta}\}^2}\right]f(x)dx=\frac{1}{6}$$

$$\int_0^\infty\left[\frac{(x/\phi)^{2\beta}}{\{1+(x/\phi)^{\beta}\}^2}\right]f(x)dx=\frac{1}{3}$$

$$\int_0^\infty\left[\frac{(x/\phi)^{\beta}\ln(x/\phi)}{\{1+(x/\phi)^{\beta}\}^2}\right]f(x)dx=-\frac{\phi}{6\beta}[\ln(\beta/\phi)]$$

and

$$\int_0^\infty \left[\frac{(x/\phi)^\beta \ln(x/\phi)^2}{\{1+(x/\phi)^\beta\}^2} \right] f(x)dx = \frac{1}{\beta^2} \frac{(\pi^2-6)}{18}.$$

The log likelihood is

$$\ln L = r\ln(1-\alpha) + (n-r)\left[\ln\alpha + \ln\beta - \beta\ln\phi\right] + (\beta-1)\sum_{x_i>0}\ln x_i - 2\sum_{x_i>0}\ln\left[1+\frac{x_i}{\phi}\right]^\beta$$

and the Fisher information's are

$$I_{\alpha\alpha} = E\left(-\frac{\partial^2 \ln L}{\partial\alpha^2}\right) = \frac{1}{\alpha(1-\alpha)}$$

$$I_{\alpha\beta} = I_{\beta\alpha} = E\left(-\frac{\partial^2 \ln L}{\partial\alpha\partial\beta}\right) = 0$$

$$I_{\beta\beta} = E\left(-\frac{\partial^2 \ln L}{\partial\beta^2}\right) = \frac{\alpha}{\beta^2}\left[1+\frac{(\pi^2-6)}{9}\right]$$

$$I_{\alpha\phi} = I_{\phi\alpha} = E\left(-\frac{\partial^2 \ln L}{\partial\alpha\partial\phi}\right) = 0$$

$$I_{\phi\beta} = I_{\beta\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\phi\partial\beta}\right) = \frac{\alpha}{\phi} - \frac{\alpha}{3\phi}\left[4-\phi\ln\left(\frac{\beta}{\phi}\right)\right]$$

and

$$I_{\phi\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\phi^2}\right) = \frac{\alpha\beta^2}{3\phi^2}$$

The above computations for both criteria's are done for Vannman's example given in section (2.8).

2.2.3 Uniformly Minimum Variance Unbiased Estimator (UMVUE)

One can obtain UMVUE of mixture density of instantaneous and positive observation taken from Pareto distribution using the method discussed in Singh (2007). Based on above families we define a new family of df $\Im = \{F(x;\theta,\alpha) : x \ge 0, \theta \in \Omega, 0 < \alpha < 1\}$ such that

$$f(x;\theta,\alpha)=\begin{cases}1-\alpha+\alpha f(x;\theta), & x=0\\ \alpha f(x;\theta), & x>0\end{cases}$$

Hence the pdf of mixture family is obtained as

$$f(x;\theta,\alpha)=(1-\alpha)^{\rho}\,\alpha^{(1-\rho)}\left[\frac{1}{(1+x)}\frac{\theta}{(1+x)^{\theta}}\right]^{(1-\rho)}$$

$$=\frac{\left(\frac{1}{1+x}\right)^{(1-\rho)}\left[\exp(-\theta)\right]^{(1-\rho)\ln(1+x)}\left(\frac{1-\alpha}{\alpha\theta}\right)^{\rho}}{\left(\frac{1}{\alpha\theta}\right)} \tag{2.2.8}$$

which is a member of exponential family with $a(x)=\left(\frac{1}{1+x}\right)$, $h(\theta)=\exp(-\theta)$, $g(\theta)=\frac{1}{\theta}$ and $d(x)=\ln(1+x)$. We have $z=\sum_{x>0}(1-\rho)\ln(1+x)$ and $n-r=\sum_{x>0}\rho_j$ which are jointly complete sufficient statistics for (θ,p). Since $\ln(1+x)$ has exponential distribution with parameter θ.

The UMVUE of mixture density is given by

$$\zeta_x(z,r,n)=\begin{cases}\dfrac{B(z,r,n-1)}{B(z,r,n)}=\dfrac{n-r}{n}, & x=0,\ r=0,1,2....n-1\\[2ex] a(x)\dfrac{B(z-d(x),r-1,n-1)}{B(z,r,n)}, & x>0,\ z>d(x),\ r=1,2....n\end{cases}$$

where

$$B(z,r,n)=\begin{cases}\binom{n}{r}B(z\,|\,r), & z=r,r+1.....;\ r=1,2....n\\ 1, & z=0,\ r=0\end{cases}$$

and $B(z\,|\,r)$ is such that

$$\left[b(\theta)-a(0)\right]^{r}=\sum_{z=r}^{\infty}B(z\,|\,r)\theta^{r},\ r=1,2,....n.$$

The above expression simplifies to

$$\zeta_x(z,r,n)=\begin{cases}\dfrac{n-r}{n}, & x=0\\ \dfrac{r(r-1)}{nzx}\left[1-\dfrac{\ln(1+x)}{z}\right]^{r-2}, & x>0\end{cases} \tag{2.2.9}$$

which is UMVUE of mixture density of instantaneous failure and positive observation taken from Pareto distribution.

2.3 Analysis for early failures

If early failures are nominally reported as $X=\delta$ then the df of the modified model G_1 is given as

$$G_1(x,\alpha,\theta)=\begin{cases}0, & x<\delta\\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta\\ 1-\alpha+\alpha F(x,\theta), & x>\delta\end{cases} \tag{2.3.1}$$

The corresponding pdf is given as

$$g_1(x,\alpha,\theta)=\begin{cases}0, & x<\delta\\ 1-\alpha+\alpha F(\delta,\theta), & x=\delta\\ \alpha f(x,\theta), & x>\delta\end{cases} \tag{2.3.2}$$

The Fisher informations can be obtained as

$$I_{\alpha\alpha} = \frac{1-F(\delta,\theta)}{\alpha\left[1-\alpha+\alpha F(\delta,\theta)\right]} \tag{2.3.3}$$

$$I_{\theta\theta} = \alpha\left[I_f(\theta)-\int_0^{\delta}\left(\frac{\partial \ln f}{\partial\theta}\right)^2 f(x,\theta)dx\right] \tag{2.3.4}$$

and

$$I_{\alpha\theta}=I_{\theta\alpha}=\frac{-\dfrac{\partial}{\partial\theta}F(\delta,\theta)}{\left[1-\alpha+\alpha F(\delta,\theta)\right]} \tag{2.3.5}$$

where $I_f(\theta)$ is the Fisher information about θ in the original pdf $f(x,\theta)$. Again using (2.3.4), we get the Fisher information about θ ignoring α as

$$I_{g_1}{}^{(\alpha)}(\theta) = \frac{\alpha\left[I_f(\theta) - \int_0^{\delta}\left(\frac{\partial \ln f}{\partial \theta}\right)^2 f(x,\theta)dx - \left(\frac{\partial F(\delta,\theta)}{\partial \theta}\right)^2\right]}{1-\alpha+\alpha F(\delta,\theta)\left[1-F(\delta,\theta)\right]} \tag{2.3.6}$$

Here one can see that the parameters α and θ are not orthogonal. Also as $0 < \alpha < 1$, $I_{g_1}{}^{(\alpha)}(\theta) < I_f(\theta)$. If the *n* observations $X_1, X_2, ..., X_n$ are from $g_1 \in \mathcal{G}_1$, then the likelihood is

$$L(x,\alpha,\theta) = \left[1-\alpha+\alpha F(\delta,\theta)\right]^{n_0} \alpha^{n-n_0} \prod_{x_i>\delta} f(x_i,\theta)\,.$$

Then the ML estimates are the solutions of the following likelihood equations:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{-n_0\left[1-F(\delta,\theta)\right]}{1-\alpha+\alpha F(\delta,\theta)} + \frac{n-n_0}{\alpha} = 0 \tag{2.3.7}$$

and

$$\frac{\partial \ln L}{\partial \theta} = \frac{-n_0 \dfrac{\partial}{\partial \theta} F(\delta,\theta)}{1-F(\delta,\theta)} + \sum_{x_i>\delta} \frac{\partial \ln f(x_i,\theta)}{\partial \theta} = 0 \tag{2.3.8}$$

Equation (2.3.8) does not depend on α and hence one can obtain $\hat{\theta}$ from (2.3.8). Using this $\hat{\theta}$ in (2.3.7) we can obtain $\hat{\alpha}$. Again,

$$L(x,\alpha,\theta) = \left[1-\alpha+\alpha F(\delta,\theta)\right]^{n_0} \left(\alpha\left[1-F(\delta,\theta)\right]\right)^{n-n_0} \prod_{x_i>\delta} \frac{f(x_i,\theta)}{1-F(\delta,\theta)}$$

That is, the likelihood of the sample under $g_1 \in \mathcal{G}_1$ is the product of the likelihoods of n_0 and the conditional likelihood of the sample given n_0 which is same as the likelihood of $(n-n_0)$ observations coming from the truncated version of $f \in \mathfrak{I}$ or ($g_1 \in \mathcal{G}_1$) restricted to (δ,∞). Now n_0 is binomial with probability of success given by $1-\alpha+\alpha F(\delta,\theta)$. For fixed θ and $\alpha \in [0,1]$ this binomial family is complete. Therefore, the optimal estimating equation for θ ignoring α is the conditional score function given n_0

or $\dfrac{\partial \ln L_{n_0}}{\partial \theta} = 0$, where $L_{n_0} = \prod_{x_i > \delta} \dfrac{f(x_i, \theta)}{1 - F(\delta, \theta)}$. Hence optimal estimating equation for θ ignoring α is given by (2.3.8). Thus $\dfrac{\partial \ln L_{n_0}}{\partial \theta}$ or $\hat{\theta}$ is same as the estimator given by optimal estimating equation for θ ignoring α.

For some computations of one parameter Pareto family defined in section (2.1), we have the following formulas:

$$\int_{\delta}^{\infty} \left[\frac{x^{\beta} \ln x^2}{\{1 + x^{\beta}\}^2} \right] f(x) dx = \frac{1 + 3u + (1 + \ln u)(2\ln u + 6u \ln u)}{12(1+u)^3}, \; u = \delta^{\beta}.$$

The log likelihood of early failure in one parameter Pareto model is

$$\ln L = r \ln\left[1 - \frac{\alpha}{(1 + \delta^{\beta})} \right] + (n - r)[\ln \alpha + \ln \beta] + (\beta - 1) \sum_{x_i > \delta} \ln x_i - 2 \sum_{x_i > \delta} \ln(1 + x_i{}^{\beta})$$

and the Fisher Information's are

$$I_{\alpha\alpha} = E\left(-\frac{\partial^2 \ln L}{\partial \alpha^2} \right) = \frac{1}{\alpha(1 - \alpha + u)},$$

$$I_{\alpha\beta} = I_{\beta\alpha} = E\left(-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta} \right) = -\frac{u \ln \delta}{(1+u)(1 - \alpha + u)},$$

and

$$I_{\beta\beta} = E\left(-\frac{\partial^2 \ln L}{\partial \beta^2} \right) = \frac{\alpha}{\beta^2} \left[\frac{1}{1+u} + \left\{ \frac{1 + 3u + (1 + \ln u)(2\ln u + 6u \ln u)}{6(1+u)^3} \right\} \right]$$
$$- \frac{\alpha u (\ln \delta)^2}{(1+u)^3 (1 - \alpha + u)} \left[(1+u)(1 - \alpha + u) - u(2 + 2u - \alpha \right],$$

where $u = \delta^{\beta}$.

For two parameters Pareto family as defined in section (2.1) we have to use the following formulas:

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{\beta}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=-\frac{1}{3[1+v]^{3}}+\frac{1}{2[1+v]^{2}},$$

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{2\beta}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=\frac{1}{[1+v]}-\frac{1}{[1+v]^{2}}+\frac{1}{3[1+v]^{3}},$$

and

$$\int_{\delta}^{\infty}\left[\frac{(x/\phi)^{\beta}\ln(x/\phi)^{2}}{\{1+(x/\phi)^{\beta}\}^{2}}\right]f(x)dx=\frac{1}{6\beta}\left[-\frac{v(1+v)+v(3+v)\ln v}{(1+v)^{3}}+\ln(1+v)\right],$$

where $v=(\delta/\phi)^{\beta}$.

The log likelihood of early failures in two parameters Pareto model is

$$\ln L= r\ln\left[1-\frac{\alpha}{1+(\delta/\phi)^{\beta}}\right]+(n-r)\left[\ln\alpha+\ln\beta-\beta\ln\phi\right]+(\beta-1)\sum_{x_i>\delta}\ln x_i$$
$$-2\sum_{x_i>\delta}\ln\left[1+(x_i/\phi)^{\beta}\right]$$

Then Fisher information equations corresponding to two parameters Pareto models are as given below

$$I_{\alpha\alpha}=E\left(-\frac{\partial^{2}\ln L}{\partial\alpha^{2}}\right)=\frac{1}{\alpha(1-\alpha+v)}$$

$$I_{\alpha\beta}=I_{\beta\alpha}=E\left(-\frac{\partial^{2}\ln L}{\partial\alpha\partial\beta}\right)=-\frac{v\ln(\delta/\phi)}{(1+v)(1-\alpha+v)}$$

$$I_{\beta\beta}=E\left(-\frac{\partial^{2}\ln L}{\partial\beta^{2}}\right)=\frac{1}{\beta^{2}}\left[\frac{1}{1+v}+\left\{\frac{1+3v+(1+\ln v)(2\ln v+6v\ln v)}{6(1+v)^{3}}\right\}\right]$$
$$-\frac{\alpha v\left[\ln(\delta/\phi)\right]^{2}}{(1+v)^{3}(1+v-\alpha)}\left[(1+v-\alpha)(1+v)-v(2+2v-\alpha)\right]$$

$$I_{\alpha\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\alpha\partial\phi}\right) = \frac{\beta v}{\phi(1+v)^2}$$

$$I_{\beta\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\beta\partial\phi}\right) = \frac{1}{3\phi}\left[\frac{v(1+v+v(3+v)\ln v)}{(1+v)^3} - \frac{3(2v^2+3v+1)}{(1+v)^3} - \ln(1+v)\right]$$

$$+\frac{\alpha v[(1+v-\alpha)(1+v)\{1+\beta\ln(\delta/\phi)\} - \beta v\ln(\delta/\phi)(2+2v-\alpha)]}{\phi(1+v-\alpha)(1+v)^3} - \frac{1}{\phi(1+v)}$$

and

$$I_{\phi\phi} = E\left(-\frac{\partial^2 \ln L}{\partial\phi^2}\right) = \frac{\beta}{\phi^2}\left[-\frac{1}{(1+v)} + \frac{(2+v)(\beta+1)}{6(1+v)^3} + \frac{2\{3(1+v^2)+5v\}}{3(1+v)^3}\right.$$

$$\left.+\frac{2\alpha\beta v\{(1+v-\alpha)(1+v)(1+\beta) - \beta v(2+2v-\alpha)}{\phi^2(1+v-\alpha)(1+v)^3}\right],$$

where $v = (\delta/\phi)^{\beta}$.

The above information is used in illustration given in section (2.8) with comparative study of instantaneous failures and early failures are presented for different situations.

2.4 Nearly instantaneous model

As already discussed in chapter 1 nearly instantaneous model incorporates inliers in better way than the above two models.

2.4.1 Representation of the model

Let $F(x)$ and $R(x)=1-F(x)$ denote the cumulative distribution function and the survival function of the mixture, respectively. We assume that F is continuous and its density be given by $f(x)=F'(x)$. The component distribution functions and their survival functions are $F_i(x)$ and $R_i(x)=1-F_i(x)$ respectively, i=1, 2. The failure rate of a lifetime distribution is defined as $h(x)=f(x)/R(x)$ provided the density exists.

We now represent this model as a mixture of the generalized Dirac delta function and the 2-parameter Pareto as opposed to a mixture of a singular distribution with Pareto, as

$$f(x) = p\delta_d(x - x_0) + q\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(x/\beta\right)^{\alpha}\right]^{-2} \quad ,p+q=1, 0<p<1 \;\; ,x>0 \quad ,\alpha>0,\beta>0 \tag{2.4.1}$$

where

$$\delta_d(x-x_0) = \begin{cases} \frac{1}{d}, & x_0 \le x \le x_0+d \\ 0, & otherwise \end{cases} \tag{2.4.2}$$

for sufficiently small d. Here $p > 0$ is the mixing proportion. Also note that

$$\delta(x - x_0) = \lim_{d\to 0} \delta_d(x - x_0) \tag{2.4.3}$$

where $\delta(\bullet)$ is the Dirac delta function. We may view the Dirac delta function as approximately normal distribution having a zero mean and standard deviation that tends to 1. For fixed value of d, equation (2.4.2) denotes a uniform distribution over an interval $[x_0, x_0+d]$ so the modified model is now effectively a mixture of a Pareto with a uniform distribution. Instead of including a possible instantaneous failure in the model (2.4.2) is allowed for a possible “nearly instantaneous” failure to occur uniformly over a very small time interval. Note that the case $x_0 = 0$ corresponds to instantaneous failures, whereas $x_0 \neq 0$ (but small) corresponds to the case with early failures. Noting from (2.4.1) and (2.4.2), we see that the mixture density function is not continuous at x_0 and x_0+d. However, both the distribution and survival functions are continuous. Writing

$$f_1(x) = \delta_d(x-x_0) \text{ and } f_2(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(x/\beta\right)^{\alpha}\right]^{-2} \quad \alpha,\beta>0$$

can be written as

$$f(x) = pf_1(x) + qf_2(x) \quad \text{where } p+q=1,\; 0<p<1 \tag{2.4.4}$$

So

$$F(x) = pF_1(x) + qF_2(x) \tag{2.4.5}$$

and

$$R(x) = 1 - F(x) = p + q - \left[pF_1(x) + q\,F_2(x)\right] = pR_1(x) + qR_2(x) \tag{2.4.6}$$

Thus, the failure (hazard) rate function of the mixture distribution is

$$h(x) = \frac{pf_1(x) + qf_2(x)}{pR_1(x) + qR_2(x)} \tag{2.4.7}$$

A mixture distribution involving two 2-parameter Weibull distribution has been thoroughly studied by Lai, Khoo, Murlidharan (2007). The mixture considered was more complex in the sense that one of the mixing distributions has a finite range which poses some challenges. Simulated observations from this model are made by generating uniform variates and Pareto variates with proportions p and $q = 1 - p$ respectively.

2.4.2 Survival function, failure rate and mean residual life function of the nearly instantaneous model

Recently, failure rates of mixtures are discussed quite extensively. The Reliability (survival) functions of the respective component distributions are given by

$$R_1(x) = \begin{cases} 1, & 0 \le x < x_0 \\ \dfrac{d + x_0 - x}{d}, & x_0 \le x \le x_0 + d \\ 0, & x \ge x_0 + d \end{cases} \tag{2.4.8}$$

and

$$R_2(x) = \frac{1}{1 + \left(x/\beta\right)^{\alpha}} \tag{2.4.9}$$

The failure rates are, respectively,

$$h_1(x) = \begin{cases} 0, & 0 \le x < x_0 \\ \dfrac{1}{d + x_0 - x}, & x_0 \le x \le x_0 + d \\ \infty, & x \ge x_0 + d \end{cases} \tag{2.4.10}$$

and

$$h_2(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\left[1 + \left(x/\beta\right)^{\alpha}\right]^{-1} \tag{2.4.11}$$

It can be shown (2.4.4) and (2.4.6) that for any mixture of two continuous distributions the failure rate function can be expressed as

$$h(x) = \frac{f(x)}{R(x)} = w(x)h_1(x) + [1 - w(x)]h_2(x) \tag{2.4.12}$$

where $w(x) = p\,R_1(x)/R(x)$ for all $x \ge 0$. In our case,

$$w(x) = \begin{cases} \dfrac{p}{R(x)}, & 0 \le x < x_0 \\ \dfrac{pR_1(x)}{R(x)}, & x_0 \le x \le x_0 + d \\ 0, & x \ge x_0 + d \end{cases} \tag{2.4.13}$$

with

$$w'(x) = w(x)\left[1 - w(x)\right]\left\{h_2(x) - h_1(x)\right\} \tag{2.4.14}$$

Also a simple differentiation shows that

$$h'(x) = w'(x)h_1(x) + w(x)h_1'(x) - w'(x)h_2(x) + \left[1 - w(x)\right]h_2'(x) \tag{2.4.15}$$

Now $w(x)\,h_1(x) = \dfrac{pR_1(x)}{R(x)}\dfrac{f_1(x)}{R_1(x)} = \dfrac{pf_1(x)}{R(x)}$, so (2.4.12) is well defined for all $x > 0$. Summarized expression for $R(x), h(x)$ and $m(x)$ are, respectively, given as

$$R(x) = pR_1(x) + qR_2(x)$$

$$R(x)=\begin{cases} p+q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & 0\le x<x_0 \\ \dfrac{p[d+x_0-x]}{d}+q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x_0\le x\le x_0+d \\ q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d \end{cases} \tag{2.4.16}$$

Recall that *h*(*x*) is discontinuous at both $x=x_0$ and $x=x_0+d$.

$$h(x)=\begin{cases} \left[\dfrac{q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}}{p+q\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}}\right]\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & 0\le x\le x_0 \\ \dfrac{p+dq\alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-2}}{p(d-x)+dq\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}}, & x_0\le x\le x_0+d \\ \alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d \end{cases} \tag{2.4.17}$$

Then the Mean residual life of an r.v. X is defined for all *x* as

$$m_X(x)=E(X-x/X>x)=\frac{\int_x^{\infty} R_X(y)dy}{R_X(y)}$$

This is the expected additional time to failure given survival to *x*.

$$m(x)=pm_1(x)+qm_2(x) \tag{2.4.18}$$

where

$$m_1(x)=\begin{cases}\dfrac{x_0-x}{2}, & 0\le x<x_0\\ \dfrac{x_0+d-x}{2}, & x_0\le x<x_0+d\\ 0, & x>x_0+d\end{cases}\tag{2.4.19}$$

$$m_2(x)=\frac{\displaystyle\int_x^\infty \frac{1}{\left[1+\left(y/\beta\right)^\alpha\right]}dy}{\dfrac{1}{\left[1+\left(x/\beta\right)^\alpha\right]}},\qquad y>x_0+d\tag{2.4.20}$$

2.4.3 Nearly instantaneous failure case ($x_0 = 0$)

Consider a special case of model (2.4.1) whereby $x_0 = 0$. The model may be called the Pareto with "nearly instantaneous failure" model. In this case, (2.4.10) is simplified giving the failure rate of the uniform distribution as

$$h_1(x)=\begin{cases}\dfrac{1}{d-x}, & o\le x\le d\\ \infty, & x>d\end{cases}\tag{2.4.21}$$

and corresponding to (2.4.8) its survival rate function is given as

$$R_1(x)=\begin{cases}\dfrac{d-x}{d}, & o\le x\le d\\ 0, & x>d\end{cases}\tag{2.4.22}$$

The Pareto model with "nearly instantaneous failure" occurring uniformly over [0, d] has

$$R(x)=\begin{cases}\dfrac{p(d-x)}{d}+q\left[1+\left(\dfrac{x}{\beta}\right)^\alpha\right]^{-1}, & 0\le x\le d\\ q\left[1+\left(\dfrac{x}{\beta}\right)^\alpha\right]^{-1}, & x>d\end{cases}\tag{2.4.23}$$

and

$$h(x)=\begin{cases}\dfrac{p+dq\alpha\beta^{\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-2}}{p(d-x)+dq\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}}, & 0\le x\le x_0+d\\ \alpha\beta^{-\alpha}x^{\alpha-1}\left[1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right]^{-1}, & x>x_0+d\end{cases} \tag{2.4.24}$$

We now present some graphical plots of Survival, Density and Failure Rate Functions. Graphical plots are important for ageing distributions. Some graphs are plotted to identify whether the model is useful for specific datasets for which empirical plots are available. All plots are done when $x_0 = 0$, the Pareto with "nearly" instantaneous failure model. A plot of density function, Survival function and MRL functions for various values of p are given below.

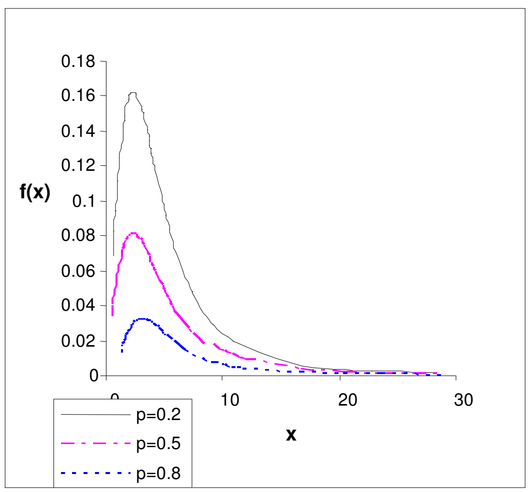


Fig. 2.4.1. Density function f(x): β =3, α =2, d = 0.5, x_0 = 0.

Failure Rate Functions. The failure rate function is given in fig. (2.4.7). Clearly, its shape is the same as the Pareto distribution after d. Thus we focus on the segment from 0 to d. The following four figures show that $h(x)$ can be increasing, decreasing, or bathtub shaped for $0 \le x \le d$. From the plots, it can be seen that the failure rate function of the model gives rise to several different shapes and bumps; this is expected as mixing with a component distribution that has a finite range often cause some problems. Although the second part can be either increasing or decreasing, the first segment can achieve various shapes. This finding agrees with Block (2003).

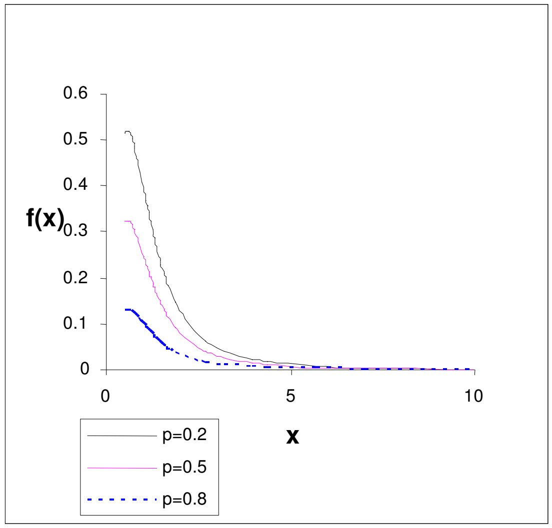


Figure 2.4.2 Density function *f(x): β* = 1,*α* =2, *d* = 0.2, x_0 = 0.

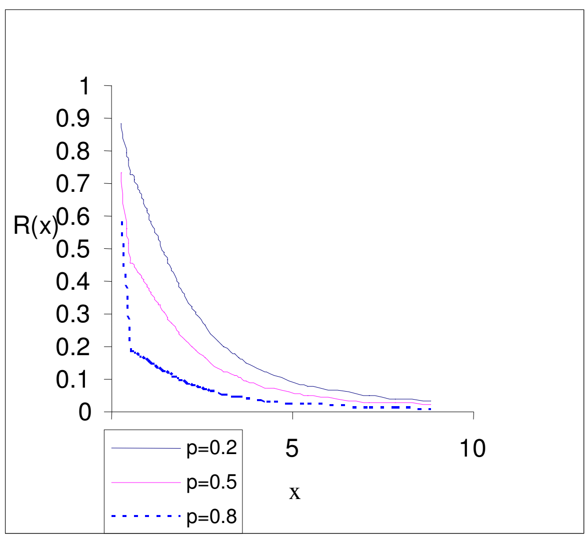


Figure 2.4.3 Reliability function R(x): *β* =3, *α* =2, *d* = 0.5, x_0 = 0

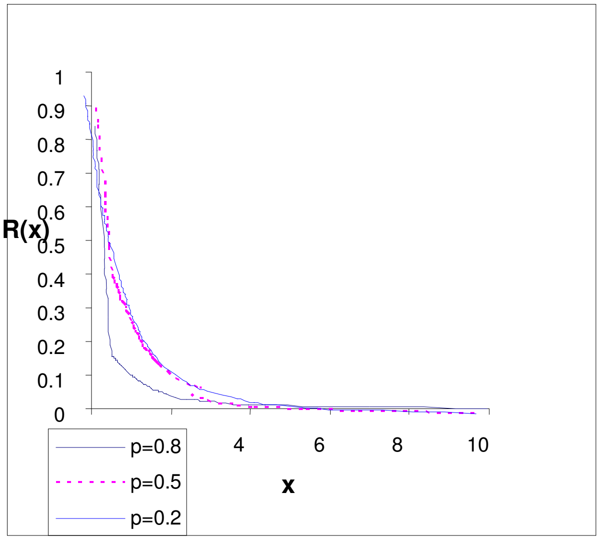


Fig. 2.4.4. Reliability function *R(x)*: β =1, α =2, d = 0.5, x_0 = 0

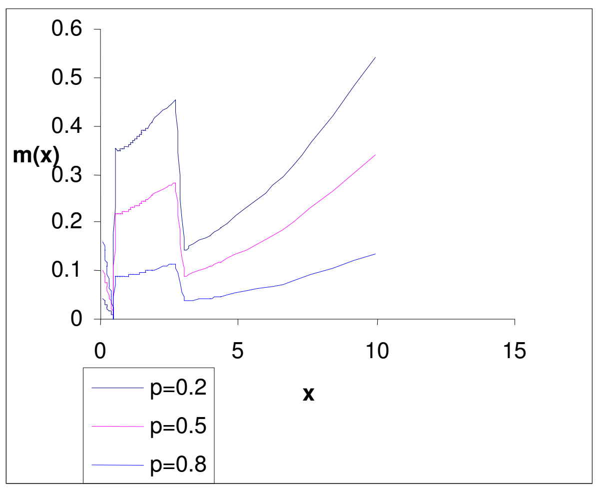


Fig. 2.4.5. Plot of mean residual *m(x)*: β =1, α =2, d = 0.5, x_0 = 0

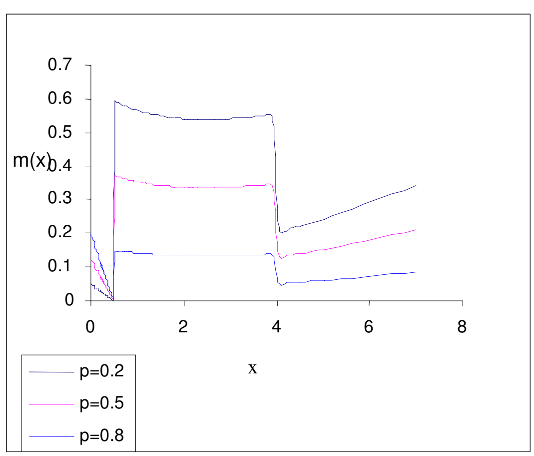
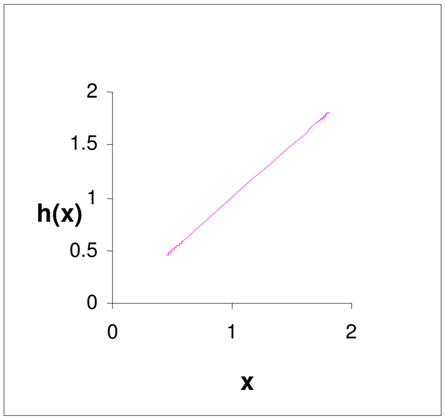
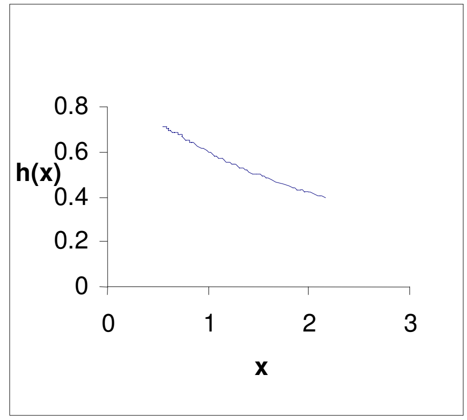


Fig. 2.4.6. Plot of mean residual m(x): β =3,α =2, d = 0.5, t_0 = 0.

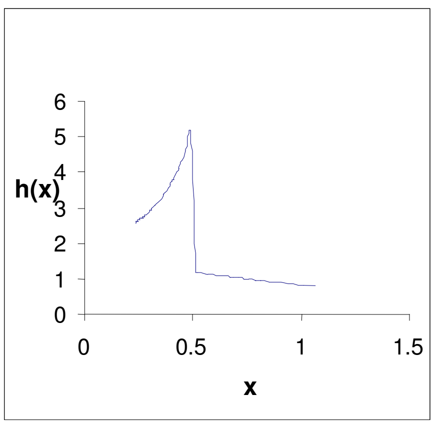
Figure (2.4.6) represents failure rate $h(x)$ for different combinations of α, β, p and d.



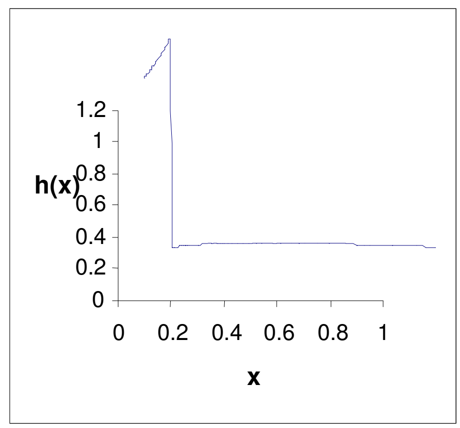
(a) α=1.2, β=1.2, p=0.3, d=0.5



(b) α=1.2, β=1, p=0.08, d=0.5



(c) α=1.2, β=0.5, p=0.5, d=0.5



(d) α=2, β=3, p=0.2, d=0.2

Fig. 2.4.7. Failure rate

2.5 Inlier estimation using L_k and M_k models

In this section we consider the situation where instantaneous (i.e. X = 0) failures can also occur by mixing a singular distribution at X = 0 with the above model of inliers. Assuming that the data is usually consisting of r_0 instantaneous failures, r_1 early failures as indicated by sample configuration and the rest $n - r_0 - r_1$ observations belong to the target population.

2.5.1 Inlier estimation for labeled slippage (L_k) models

For this model consider observations from inlier with pdf

$$g(x)=\frac{\phi}{(1+x)^{\phi+1}},x>0,\ \phi>0$$

and those from target population with pdf

$$f(x)=\frac{\theta}{(1+x)^{\theta+1}},\ x>0,\ \theta>0$$

Then the likelihood of the sample from population with observations from inliers with pdf $g\left(x_{(i)},\phi\right)$ and target pdf $f\left(x_{(i)},\theta\right)$

$$L=\binom{n}{r_0}(1-\alpha)^{r_0}\,\alpha^{n-r_0}\left[\frac{r_1!(n-r_0-r_1)!}{\varphi_{r_1}(G,F)}\right]\prod_{i=1}^{r_1}g\left(x_{(i)},\phi\right)\prod_{i=r_1+1}^{n}f\left(x_i,\theta\right) \qquad (2.5.1)$$

where $\varphi_r(G,F)$ is defined as

$$\varphi_r(G,F)=\mathrm{P}\left(X_{(r)}<X_{(r+1)}\,|\,G,\ F\right)$$

$$=\int_{-\infty}^{\infty}[G(u)]^r(n-r)[1-F(u)]^{n-r-1}dF(u) \qquad (2.5.2)$$

The likelihood function in (2.5.1) assumes that between the experiments when units are placed on test we do not know which of the units fail instantaneously. Equivalently $X_{i_1}=0,\ X_{i_2}=0,\ \ldots X_{i_{r_0}}=0$ which fail early i.e. those units whose failure time distribution is $g(x_{(i)},\phi)$ with failure rate much larger than that of the failure time distribution of the target population whose failure rate is considerably smaller. The log likelihood of the model is

$$\ln L=r_0\ln(1-\alpha)+(n-r_0)\ln\alpha-\ln\varphi_{r_1}(\phi,\theta)+r_1\ln\phi-(\phi+1)\sum_{i=1}^{r_1}\ln(1+x_{(i)})$$
$$+(n-r_0-r_1)\ln\theta-(\theta+1)\sum_{i=r_1+1}^{n-r_0}\ln(1+x_{(i)})$$

and the likelihood equations are

$$\frac{\partial \ln L}{\partial \alpha}=\frac{-r_0}{(1-\alpha)}+\frac{(n-r_0)}{\alpha}=0 \tag{2.5.3}$$

$$\frac{\partial \ln L}{\partial \phi}=-\frac{\partial}{\partial \phi}\ln\varphi_{r_1}(\phi,\theta)+\frac{r_1}{\phi}-\sum_{i=1}^{r_1}\ln\left(1+x_{(i)}\right) \tag{2.5.4}$$

and

$$\frac{\partial \ln L}{\partial \theta}=-\frac{\partial}{\partial \theta}\ln\varphi_{r_1}(\phi,\theta)+\frac{n-r_0-r_1}{\theta}-\sum_{i=r_1+1}^{n}\ln\left(1+x_{(i)}\right) \tag{2.5.5}$$

Here (2.5.3) can be solved to get the estimate of α as $\hat{\alpha}=(n-r_0)/n$. Solving (2.5.4) and (2.5.5) simultaneously we get the estimate of ϕ and θ. The parameter α is orthogonal to (ϕ,θ). The second order derivatives are

$$\frac{\partial^2 \ln L}{\partial \phi^2}=-\frac{\partial^2}{\partial \phi^2}\ln\varphi_{r_1}(\phi,\theta)-\frac{r_1}{\phi^2} \tag{2.5.6}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2}=-\frac{\partial^2}{\partial \theta^2}\ln\varphi_{r_1}(\phi,\theta)-\frac{(n-r_0-r_1)}{\theta^2} \tag{2.5.7}$$

and

$$\frac{\partial^2 \ln L}{\partial \phi\partial \theta}=-\frac{\partial^2}{\partial \phi\partial \theta}\ln\varphi_{r_1}(\phi,\theta) \tag{2.5.8}$$

where

$$\varphi_{r_1}(\phi,\theta)=(n-r_0-r_1)\int_0^\infty\left\{1-\frac{1}{(1+x)^{\phi}}\right\}^{r_1}\left[\frac{1}{(1+x)^{\theta}}\right]^{n-r_0-r_1-1}\frac{\theta}{(1+x)^{\theta+1}}dx$$

$$=\frac{(n-r_0-r_1)\theta}{\phi}\beta\left(r_1+1,\frac{(n-r_0-r_1)\theta}{\phi}\right)$$

$$=\left[\frac{(n-r_0-r_1)\theta}{\phi}\right]\left[\frac{\Gamma(r_1+1)\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}\right)}{\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}+r_1+1\right)}\right] \tag{2.5.9}$$

Taking log on both sides we get

$$\ln\varphi_{r_1}(\phi,\theta)=C+\ln\theta-\ln\phi+\ln\Gamma(z)-\ln\Gamma((z+r_1+1))$$

where

$$z = \{(\mathrm{n} - \mathrm{r}_0 - \mathrm{r}_1)\theta\} / \phi$$

and

$$\frac{\partial}{\partial\phi}\ln\varphi_{r_1}(\phi,\theta) = -\frac{1}{\phi} + \frac{\partial}{\partial\phi}\ln\Gamma z\frac{\partial z}{\partial\phi} - \frac{\partial}{\partial\phi}\ln\Gamma(z + r_1 + 1)\frac{\partial z}{\partial\phi}$$

$$= -\frac{1}{\phi} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(-\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)$$

where

$$\psi(z) = \frac{\partial}{\partial\phi}\Gamma z \quad and\, \Gamma z = \int_0^\infty x^{z-1}e^{-x}dx$$

The second derivative of the likelihood functions are

$$\frac{\partial^2 \ln\varphi_{r_1}(\phi,\theta)}{\partial\phi^2} = \frac{1}{\phi^2} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{2(n - r_0 - r_1)\theta}{\phi^3}\right)$$
$$-\left\{\frac{(n - r_0 - r_1)\theta}{\phi^2}\left[\psi'(z) - \psi'(z + r_1 + 1)\right]\left(-\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)\right\}$$

$$\frac{\partial^2 \ln\varphi_{r_1}(\phi,\theta)}{\partial\phi^2} = \frac{1}{\phi^2} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{2\theta(n - r_0 - r_1)}{\phi^3}\right)$$
$$+\left\{\left(\frac{(n - r_0 - r_1)\theta}{\phi^2}\right)^2\left[\psi'(z) - \psi'(z + r_1 + 1)\right]\right\}$$

where $\quad \psi'(z) = \dfrac{\partial^2}{\partial\phi^2}\ln\Gamma z$

now

$$\frac{\partial}{\partial\theta}\ln\varphi_{r_1}(\phi,\theta) = \frac{1}{\theta} + \frac{\partial}{\partial\theta}\ln\Gamma z\frac{\partial z}{\partial\theta} - \frac{\partial}{\partial\theta}\ln\Gamma(z + r_1 + 1)\frac{\partial z}{\partial\theta}$$

$$= -\frac{1}{\theta} + \left[\psi(z) - \psi(z + r_1 + 1)\right]\left(\frac{(n - r_0 - r_1)}{\phi}\right)$$

$$\frac{\partial^2 \ln\varphi_{r_1}(\phi,\theta)}{\partial\theta^2} = -\frac{1}{\theta^2} + \left[\psi'(z) - \psi'(z + r_1 + 1)\right]\left(\frac{(n - r_0 - r_1)}{\phi}\right)^2$$

$$\frac{\partial^2 \ln\varphi_{r_1}(\phi,\theta)}{\partial\phi\partial\theta} = \left[\psi(z)-\psi(z+r_1+1)\right]\left(-\frac{(n-r_0-r_1)}{\phi^2}\right)$$
$$-\left(\left\{\frac{(n-r_0-r_1)}{\phi}\right\}\left(\frac{(n-r_0-r_1)\theta}{\phi^2}\right)\left[\psi'(z)-\psi'(z+r_1+1)\right]\right)$$

Using results from Abramovitz and Stegun (1965) we get

$$\left[\psi(z)-\psi(z+r_1+1)\right] = -\sum_{j=1}^{r_1}\frac{1}{z+j} \tag{2.5.10}$$

$$\left[\psi'(z)-\psi'(z+r_1+1)\right] = \sum_{j=1}^{r_1}\frac{1}{(z+j)^2} \tag{2.5.11}$$

Using the above results, we obtain the likelihood equations as

$$\frac{\partial \ln L}{\partial\phi} = 0 \Rightarrow = \frac{r_1+1}{\phi} - \frac{\theta(n-r_0-r_1)}{\phi^2}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right] - \sum_{i=1}^{r_1}\ln(1+x_i) = 0 \tag{2.5.12}$$

$$\frac{\partial \ln L}{\partial\theta} = 0 \Rightarrow = \frac{(n-r_0-r_1-1)}{\theta} + \frac{(n-r_0-r_1)}{\phi}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right] - \sum_{i=r_1}^{n}\ln(1+x_i) = 0 \tag{2.5.13}$$

The above equations may be solved simultaneously to get estimates for ϕ and θ.

The Fisher information's are obtained as

$$I_{\phi\phi} = E\left[\frac{-\partial^2 \ln L}{\partial\phi^2}\right] = \frac{r_1+1}{\phi^2} - \frac{2\theta(n-r_0-r_1)}{\phi^3}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right] + \left\{\frac{(n-r_0-r_1)\theta}{\phi^2}\right\}^2 \sum_{j=1}^{r_1}\frac{1}{(z+j)^2}$$

$$I_{\theta\theta} = E\left[\frac{-\partial^2 \ln L}{\partial\theta^2}\right] = \frac{n-r_0-r_1-1}{\theta^2} + \left\{\frac{(n-r_0-r_1)}{\phi}\right\}^2 \sum_{j=1}^{r_1}\frac{1}{(z+j)^2}$$

and

$$I_{\phi\theta} = E\left[\frac{-\partial^2 \ln L}{\partial\phi\partial\theta}\right] = \frac{\theta(n-r_0-r_1)^2}{\phi^3}\left[\sum_{j=1}^{r_1}\frac{1}{(z+j)^2}\right] + \left\{\frac{(n-r_0-r_1)}{\phi^2}\right\}\sum_{j=1}^{r_1}\frac{1}{(z+j)}$$

The graph of $\varphi_r(G,F)$ to detect inlier is represented on graph (2.5.2).

2.5.2 Identified Inliers Model (M_k model)

Here we assume that the failure times $\left(X_1, X_2, \ldots\ldots X_n\right)$ of *n* units put on test are such that (*n-r*) of them are i.i.d. with FTD belonging to $\Im$ characterizing target population and remaining *r* are i.i.d. with FTD from $\mathcal{G}$ causing inlier observations where $G \in \mathcal{G}$ and $F \in \Im$ are such that $\dfrac{\partial G}{\partial F}$ is decreasing in *X*. As the indexing set *v* and the number of inliers are known we can relate $\left(X_1, X_2, \ldots\ldots X_r\right)$ i.i.d. as $G \in \mathcal{G}$ are independently distributed of $\left(X_{r+1}, X_{r+2}, \ldots\ldots X_n\right)$ from $F \in \Im$. Then the likelihood of the sample is given by

$$L\left(x \mid \phi, \theta, v, r\right) = \prod_{i=1}^{r} g\left(x_i\right) \prod_{i=r+1}^{n} f\left(x_i\right) \tag{2.5.14}$$

The MLE of parameter of *G* and *F* is a straight forward two sample problem. Suppose that the target population has FTD given by pareto distribution parameter θ and the inliers are given by pareto distribution with parameter ϕ where $\phi > \theta$ and the likelihood in the identified inlier model is given by

$$L\left(x \mid \phi, \theta, v, r\right) = \prod_{i=1}^{r} \frac{\phi}{\left(1 + x_{(i)}\right)^{\phi+1}} \prod_{i=r+1}^{n} \frac{\theta}{\left(1 + x_{(i)}\right)^{\theta+1}} \tag{2.5.15}$$

For each r =1, 2,...n we find maximum likelihood using equation (2.5.15), and then consider inlier $\hat{r}$ being that value of *r* for which likelihood is maximum.

2.5.3 Simulation study

To illustrate the method of identifying inliers model we have generated 15 independent random samples, where five of them are coming from Pareto with parameter $\phi = 20$ and remaining ten observations from Pareto distribution with parameter $\theta = 0.8$. The samples are 0.01339, 0.02679, 0.03442, 0.05519, 0.09459, 0.32854, 0.64367, 1.19427, 3.00276, 3.14612, 3.15643, 3.94635, 5.17659, 9.79405 and 12.52736. The model under illustration is identified inliers model. The identification is done as follows to evaluate for each fixed r the maximum likelihood equation $\hat{L}_r$ and then consider $\hat{r}$ being that value of r for which likelihood is maximum. The estimates have been presented in table (2.5.1).

It is interesting to note that the maximum likelihood corresponds to $\hat{r} = 5$, which was expected. The corresponding estimates of the parameters are $\hat{\phi}$ = 22.96948 and $\hat{\theta}$ =0.704261. The graphical representations of the likelihood plot are given in figure (2.5.1).

Table 2.5.1. The Likelihood and parameter estimates

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{L}_r$
1	75.18149	0.971976	8.46574E-12
2	50.32893	0.904208	1.14685E-10
3	40.77225	0.836623	1.33608E-09
4	31.42176	0.769788	9.1914E-09
5	**22.96948**	**0.704261**	**3.21716E-08**
6	12.06675	0.646565	1.02195E-08
7	7.041070	0.596001	2.28998E-09
8	4.494340	0.553932	4.46583E-10
9	2.841800	0.533335	4.66372E-11
10	2.179040	0.508762	1.37914E-11
11	1.829110	0.476013	6.59953E-12
12	1.576350	0.440886	3.38237E-12
13	1.378100	0.401308	1.74583E-12

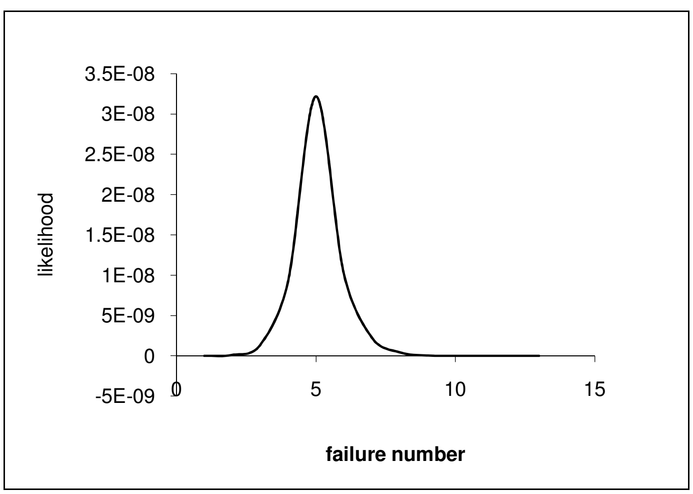


Fig. 2.5.1. Likelihood plot

Figure 2.5.2. The graph of $\varphi_r(G,F) = P[X_{(i)} < X_{(i+1)}]$

Clearly the above graph also indicates the number of inliers is 5.

2.6 Inliers detection using information criterion

The most important use of information criterion is, that it helps us in model selection, from the set of different models which all fit the data. These criteria are suitable when the underlying distribution and inlier distribution are available. It is an exploratory data analysis approach as no formal statistical inference is performed. Here three information criteria are discussed, to detect number of inliers in the data set, such as Schwarz's Information criterion ($SIC = 2\ln L(\Theta) + p\ln n$), Schawarz's Bayesian Information criterion ($BIC = -\ln L(\Theta) + \frac{0.5(p\ln n)}{n}$)and Hannan-Quinn criterion : $HQ = -\ln L(\Theta) + p\ln(\ln(n))$ where L(Θ) the maximum likelihood function and *p* is the number of free parameters that need to be estimated under the model. Below we develop the procedure for SIC scheme:

Denoting the parameter of X by α_i, *i*=1,2,...*n*. The following model of no inliers where X is from one parameter Pareto distribution with pdf

$$f(x) = \frac{\theta}{(1+x)^{\theta+1}}, x > 0, \theta > 0.$$

Now let

$$\text{Model}(0): \alpha_i = \theta \quad i = 1,2,......p \ , \tag{2.6.1}$$

And the model with *r* inliers as

$$\text{Model}(\text{r}): \alpha_i = \begin{cases} \phi, & 1 \le i < r \\ \theta, & r+1 \le i < n \end{cases} \tag{2.6.2}$$

where inliers have pdf $g(x) = \dfrac{\phi}{(1+x)^{\phi+1}}$, $x > 0$, $\phi > 0$ and *r* is such that $1 \le r \le n$, is the unknown index of the inliers. Model(0) may also be interpreted as having all observations from the target distribution *F* with common parameter .

Suppose that the life times of $X_1, X_2, ..., X_n$ is sequence of independent random variables with Pareto distribution having unknown parameter θ. Our aim is to detect those information's(inliers) from the *n* models given by equation (2.6.2).

According to the procedure, the model(0) is selected with no inliers if $SIC(0) < \min\limits_{1 \le r \le n-1} SIC(r)$. And the model(*r*) is selected if $SIC(0) > \min\limits_{1 \le r \le n-1} SIC(r)$. For Pareto distribution, the model with 0 inlier is given by

$$SIC(0) = -2n\ln\theta + 2(\theta+1)\sum_{i=1}^{n}\ln(1+x_i) + p\ln n \tag{2.6.3}$$

and

$$SIC(r) = -2r\ln\phi - 2(n-r)\ln\theta + 2(\phi+1)\sum_{i=1}^{r}\ln(1+x_i) + 2(\theta+1)\sum_{i=r+1}^{n}\ln(1+x_i) + p\ln n \tag{2.6.4}$$

where

$$\hat{\phi} = \frac{r}{\sum\limits_{i=1}^{r}\ln(1+x_i)} \quad \text{and} \quad \hat{\theta} = \frac{(n-r)}{\sum\limits_{i=r+1}^{n}\ln(1+x_i)} \tag{2.6.5}$$

The estimate of inliers say *r* is such that $SIC(r) = \min\limits_{1 \le r \le n} SIC(r)$. The illustration uses this method with the simulated example discussed in the previous section (2.5.1) and Table (2.6.1) presents the parameter estimates and the information criterion values.

Table 2.6.1. Parameter estimates and the information criterion values

$\hat{r}$	$\hat{\theta}$	$\hat{\phi}$	$SIC(r)$	$BIC(r)$	$HQ(r)$
0	1.040442	----------	60.3520	29.82228	58.64077917
1	0.971976	75.18149	53.6980	26.49499	51.98621518
2	0.904208	50.32893	48.4857	23.88883	46.77389739
3	0.836623	40.77225	43.5751	21.43353	41.86328438
4	0.769788	31.42176	39.7180	19.50500	38.00622349
5	**0.704261**	**22.96948**	**37.2124**	**18.25218**	**35.50059008**
6	0.646565	12.06675	39.50599	19.39897	37.79417263
7	0.596001	7.041079	42.4975	20.89472	40.78567634
8	0.553932	4.494343	45.7668	22.52940	44.05502113
9	0.533335	2.841807	50.2853	24.78862	48.57347388
10	0.508762	2.179042	52.7220	26.00698	51.01018362
11	0.476013	1.829118	54.1960	26.74402	52.48427275
12	0.440886	1.576359	55.5329	27.41245	53.82112047
13	0.401308	1.378105	56.8556	28.07379	55.14380810

Clearly $SIC(0)=60.3526 > SIC(5)=\min_{1\le r\le n} SIC(r)=37.21241$. A similar conclusion can be drawn in the case of other information criterions:

$$\mathrm{BIC}(0)=29.82228 > BIC(5)=\min_{1\le r\le n} BIC(r)=18.25218$$

$$HQ(0)=58.64077917 > HQ(5)=\min_{1\le r\le n} HQ(r)=35.50059008.$$

Above table clearly indicates $\hat{r}=5$ and the corresponding estimates for the parameters are $\hat{\phi}$ = 22.96948 and $\hat{\theta}$ =0.704261.

Next, we carried out an experiment with 1000 samples each of size 15 and number of inliers as 3,4,5 and 6 each with $\theta=$ 0.8 and $\phi=4,2,1.0,1.33$. The following table entitled power of SIC procedure presents the number of times the SIC procedure correctly identified the number of inliers in proportion to total number of samples. The values clearly indicate the effectiveness of the method in detecting the inliers. One of the important problem while detecting the inliers is the masking effect, where masking effect is defined as the loss of power due to wrong detection of more than one inlier.

Table 2.6.2. Power of SIC procedure

θ/ϕ \ r	0.2	0.4	0.6	0.8
3	0.055	0.083	0.098	0.103
4	0.084	0.116	0.136	0.128
5	0.102	0.153	0.158	0.157
6	0.128	0.168	0.170	0.175

2.7 Inlier estimation through Sequential Probability Ratio Test (SPRT)

To test the hypothesis whether an observation belongs to inliers population against hypothesis that it belongs to target population. The SPRT test is given as follows:

Under H_1 the pdf and likelihood function is given by

$$f(x,\theta)=\theta/(1+x)^{\theta+1}$$

and

$$L_{1m}=\prod_{i=1}^{m} f(x_i,\theta)=\prod_{i=1}^{m}\frac{\theta}{(1+x)^{\theta+1}}$$

Under H_0 the pdf and likelihood function is given by

$$g(x,\phi)=\phi/(1+x)^{\phi+1}$$

and

$$L_{0m}=\prod_{i=1}^{m} g(x_i,\phi)=\prod_{i=1}^{m}\frac{\phi}{(1+x)^{\phi+1}}$$

The likelihood ratio λ_m is given by $\lambda_m=\frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln\lambda_m=\sum_{i=1}^{m}\ln\frac{f\left(x_{(i)},\theta\right)}{g\left(x_{(i)},\phi\right)}=\sum_{i=1}^{m}z_{(i)} \qquad m=1,2,....n \tag{2.7.1}$$

For deciding number of inliers r we continue to take additional observations till we reject H_0. That is

if $\sum_{i=1}^{m}z_{(i)}\le \ln B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{m} z_{(i)} \geq \ln A$ reject H_0 and stop. The corresponding m represents the first observation from $f\left(x_{(i)},\theta\right)$ and number of inliers $r = m$-1.

$$B=\frac{\beta}{1-\alpha}\ ,\quad A=\frac{1-\beta}{\alpha} \tag{2.7.2}$$

where α represents probability of type I error and β represents probability of type II error. Hence

$$\ln\lambda_m=\sum_{i=1}^{m}\ln\frac{f\left(x_{(i)},\theta\right)}{g\left(x_{(i)},\phi\right)}=m\left(\ln\theta-\ln\phi\right)+\sum_{i=1}^{m}\ln\left(1+x_{(i)}\right)\left(\theta-\phi\right) \tag{2.7.3}$$

Arrange $X_{(1)} \leq X_{(2)} \leq ......X_{(n)}$ and apply SPRT process till the hypothesis H_0 is rejected.

Test criteria for rejection of H_0 is

$$\ln\lambda_m > \ln A=\sum_{i=1}^{m}\ln\left(1+x_{(i)}\right)>\frac{\ln A-m\left(\ln\theta-\ln\phi\right)}{\left(\theta-\phi\right)} \tag{2.7.4}$$

Corresponding value of m for which H_0 was accepted last becomes number of inliers r. The above test is conducted for the example in next section.

2.8 Illustrative Example

The main reason for detecting early failures is that the inclusion of these observations will result in underestimating life expectancy or the reliability of the item or system. This in turn may underestimate the true quality of the product. But there are situations in which instantaneous or early failures may be desirable. For example, consider the following experiment carried by Vannman (1991). A batch of wooden boards is dried by a particular chemical process and the object of the experiment is to compare two processes as regards the extent of deformation of boards due to checking. The measure of damage to the board is the checking area x defined as $x=\frac{l\bar{d}}{hl_0}100$, where l is the length of the check, $\bar{d}$ is the mean depth of the check, h is the thickness of the board area and l_0 is the length of the board. Thus *x* is the check area measured as percentage of the board area. The boards are dried at the same time under different schedule and under some climatic conditions. When drying boards not all of them will get the checks and a typical sample of wood contain several observations with x_i = 0 or x_i > 0 but relatively small compared to the rest of the checks. These observations will correspond to instantaneous failures or early failures. Note that the larger the number of instantaneous failures better is the process. Below is the reproduced data of Schedule 1 and 2 of Experiment 3 conducted by Vannman (1991). In both the case *n*=37. For data refer appendix.

First of all, we justify the Pareto model for the above data using the technique given in Meeker and Escobar(1998) and plotted *log[-log(*1-*p)]*, *p= F(*x_i*)* against *log(*x_i*)* and obtained the one parameter Pareto plot and two parameter Pareto plot separately for Schedule 2. For early failure analysis, we assumed δ=0.2. With this the observation 0.08 of Schedule 1 becomes an early failure and the observations 0.02 to 0.09 (total of 5) items of Schedule 2 become early failures.

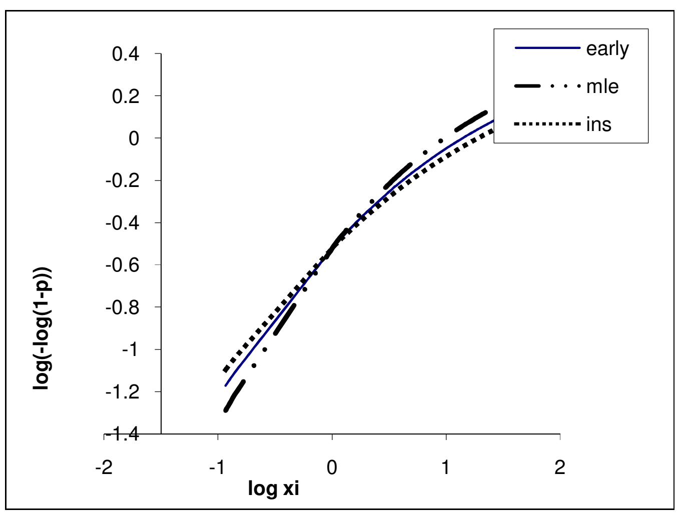


Fig. 2.8.1. One parameter Pareto plot for Schedule-2

The plots are given in figure (2.8.1) and (2.8.2) for one parameter and two parameters Pareto plot respectively for Schedule-2 of experiment 3.

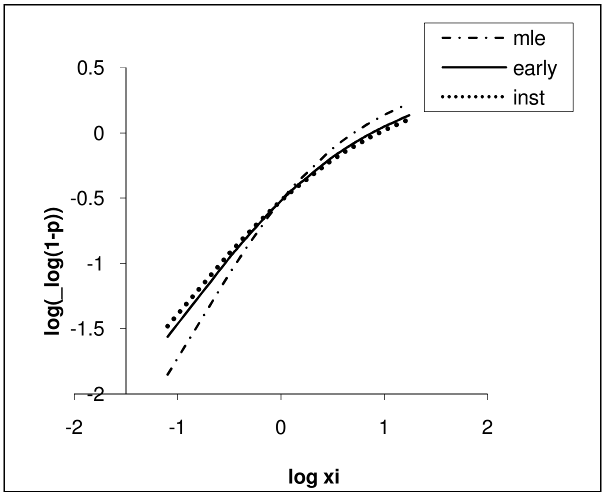


Fig. 2.8.2. Two parameter Pareto plot for Schedule-2

Table 2.8.1. Estimates of one parameter Pareto distribution

Model	Parameter	Estimates	
		Schedule-1	Schedule-2
$\mathcal{G}$ (instantaneous failures)	α	0.64865 (0.006159)	0.54054 (0.006712)
	β	1.00541 (0.029455)	0.803645 (0.022583)
$\mathcal{G}_1$ (early failures)	α	0.86603 (0.012723)	0.65394 (0.017035)
	β	0.577342 (0.011761)	0.305469 (0.003649)

Table 2.8.2. Estimates of two parameter Pareto distribution

Model	Parameter	Estimates	
		Schedule-1	Schedule-2
$\mathcal{G}$ (instantaneous failures)	α	0.64865 (0.006160)	0.54054 (0.0067123)
	β	1.25299 (0.421699)	0.800943 (0.139438)
	ϕ	2.77164 (5.638073)	0.823577 (0.393020)
$\mathcal{G}_1$ (early failures)	α	0.652919 (0.007057)	0.43217 (0.007451)
	β	1.22408 (0.034035)	1.23621 (0.036314)
	ϕ	2.30727 (0.226172)	1.79431 (0.130620)

Note: the values in the parenthesis represents variances of the estimates

The above analysis shows that the results differ in the models $\mathcal{G}$ and $\mathcal{G}_1$. In $\mathcal{G}_1$, even if we keep δ=0.1 or any value in between 0.1 to 0.2 the results are similar. Further, if we ignore the value of α then the information loss of β are 0.064116 for Schedule 1 and 0.048048 for Schedule 2 correspond to the one parameter Pareto distribution. Similarly the information loss for two parameter distributions is 0.0036226 for β and 0.00015074 for ϕ in Schedule 1 and 0.0023976 for β and 0.0004433 for ϕ in Schedule 2, respectively. Thus to retain the complete information the presence of α and δ are very much required. Moreover, from Tables 1 and 2 it is observed that the variance of the estimators of the parameters corresponds to early failures is less than the corresponding variance of instantaneous failures. Also the presence of more parameters makes the model more flexible to use. If in equation (2.1.3), the individual life times $x_i \in (0, \delta)$ are available and are not reported as δ, the problem becomes more complex.

Table 2.8.3. Estimates for instantaneous failures

Schedule		$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
I	Estimates	0.351351	1.25299	2.77164
	Standard Error	2.094713	0.508146	0.154711
II	Estimates	0.459459	0.800943	0.823577
	Standard Error	2.006607	1.02006	0.380592

Table 2.8.4. Uniform spread of "nearly instantaneous" failure times .

Schedule		$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
I	Estimates	0.378378	1.3949	3.10361
	Standard Error	2.06793	0.527328	0.159483
II	Estimates	0.567568	1.23692	1.80127
	Standard Error	2.018516	0.728331	0.298683

If we fit above data to one- parameter Pareto distribution, taking β=1, we get following estimates for the two schedules:

Table:2.8.5. Estimates for instantaneous and nearly instantaneous failure when β=1

Schedule		$\hat{p}$	$\hat{\alpha}$
I	Instantaneous	0.351351 (2.094713)	0.922071 (0.768719)
	Nearly	0.378378 (2.06193)	0.969036 (0.759074)
II	Instantaneous	0.459459 (2.0006607)	0.703233 (1.152618)
	Nearly	0.567568 (2.018516)	1.14447 (1.069799)

Note: Figures in the bracket represents the standard error of the estimates.

Table 2.8.6. Estimates of parameters and r.

r	$\hat{\theta}$	$\hat{\phi}$	$\hat{L}_r$	$SIC(r)$	$Z_{(i)}$
1	0.654913	12.99359	1.5045E-29	135.9111	0.076961
2	0.63143	5.640273	2.51883E-29	134.8804	0.354593
3	0.608352	4.433435	4.99947E-29	133.5093	0.676676
4	0.585805	3.791064	9.15284E-29	132.2999	1.055113
5	0.5654	3.141481	1.17998E-28	131.7918	1.591606
6	0.54536	2.739172	1.50475E-28	131.3056	2.190443
7	0.527292	2.368137	1.5387E-28	131.2609	2.878949
8	0.508935	2.128842	1.67179E-28	131.095	3.403319
9	**0.490793**	**1.942426**	**1.77838E-28**	**130.9714**	3.956704
10	0.479839	1.661219	1.07411E-28	131.9798	4.964562
11	0.468734	1.47416	7.37023E-29	132.7331	
12	0.459438	1.321979	4.93023E-29	133.5372	
13	0.449527	1.212013	3.60868E-29	134.1613	
14	0.441077	1.117822	2.593E-29	134.8224	
15	0.43224	1.043523	1.95208E-29	135.3902	
16	0.424204	0.979354	1.46919E-29	135.9586	
17	0.416518	0.924409	1.11832E-29	136.5043	
18	0.410398	0.874798	8.41417E-30	137.0733	
19	0.403448	0.833225	6.52924E-30	137.5806	
20	0.394893	0.797868	5.19969E-30	138.036	
21	0.385706	0.765915	4.14956E-30	138.4872	
22	0.372682	0.737522	3.35093E-30	138.9147	

For inliers detection based on section (2.5) and (2.6) we have used only schedule 1 data which are shown in table (2.8.6).Clearly, SIC(0) = 139.9487 > SIC(9) = min SIC(r) = 130.9714. Also the likelihood is maximum for r = 9. The corresponding estimates of the parameter are $\hat{\phi}$ = 1.942426 and $\hat{\theta}$ =0.490793. Using SPRT of section (2.7) the hypothesis $H_0: \phi = 2$ against $H_0: \theta = 0.5$ is also tested, for which $\alpha = 0.005$, $\beta = 0.065$. Hence *ln A* = -13.4417 and *ln B* = -2.72836 and H_0 is rejected when $Z_{(i)} = \sum_{i=1}^{m} \ln\left(1 + x_{(i)}\right) = 4.964562$ $> \frac{\ln A - m(\ln\theta - \ln\phi)}{(\theta - \phi)} = 4.340185$. SPRT also gives number of inliers as $\hat{r} = 9$.

The Pareto distribution has been used in many reliability fields. However one often finds that it does not fit well in the early part of lifespan for various reasons. In particular, in the cases where initial defects are present causing early failures, the Pareto distribution is found inadequate to model such phenomenon. The proposed model of a modified Pareto mixing with Uniform distribution to model the first phase of lifespan should provide a useful alternative.

Chapter 3

Inliers estimation in normal models

3.1 Introduction

A normal distribution is a very important statistical data distribution pattern occurring in many natural phenomena, such as height, blood pressure of person, lengths of objects produced by machines, etc. Usually normal distributions are symmetrical with a single central peak at the mean (average) of the data. But many times we may get normal distribution as mixture of inlier and target groups. For example life time of a battery follows normal distribution, it is possible in the data set, we may get two sets of observations. The first set of data may have zero or small life time compared to another group with target life time. This may create two symmetrical curved graphs, where the mean of inlier group is much less than the mean of target group. Many authors have worked on mixture of normal distributions.

In this chapter the occurrence of instantaneous or early failures in life testing experiment, which is a phenomenon observed in electronic parts as well as in clinical trials is modeled as mixture of two normal distributions. These occurrences may be due to inferior quality or faulty construction or due to no response of the treatments. The modified model is then a non-standard distribution and we call such models as inlier(s) prone models. Normal mixture distributions are arguably the most important mixture models, and also the most technically challenging. The likelihood function of the normal mixture model is unbounded based on a set of random samples, unless an artificial bound is placed on its component variance parameter. Moreover, the model is not strongly identifiable so it is hard to differentiate between over dispersion caused by the presence of a mixture and that caused by a large variance, and it has infinite Fisher information with respect to mixing proportions. There has been extensive research on finite normal mixture models, but much of it addresses merely consistency of the point estimation or useful practical procedures, and many results require undesirable restrictions on the parameter space.

In the developments below we consider $N\left(\theta,\sigma^2\right)$ as our target population, and the instantaneous and early failures are inlier components. A two parameter Normal (target) family has the probability density function

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2,\quad -\infty<x<+\infty,-\infty<\theta<+\infty,\quad \sigma>0 \qquad (3.1.1)$$

3.2 Inlier(s) prone models and estimation

Many times in real life data, we observe that data contains inliers. The data is mostly from normal population hence, we fit models which will incorporate mixture distribution of inlier and target observations with normal distributions. The assumption considered in this chapter is that the inlier and target population defer only in their mean values, where as population variances are same.

3.2.1 Normal with instantaneous failures

In a parametric model for FTD we start with a family of FTD $\Im$={F(x, θ), x≥0, $\theta \in \Omega \subset R_m\}$, where the form of the distribution function (df) is known except for labeling parameter, m-dimensional θ and F is absolutely continuous function with probability density function (pdf), $f(x,\theta)$ with respect to Lebesgue measure. The basic problem is to infer about unknown θ or a suitable function thereof say $\psi(\theta)$, on the basis of a random sample of size n on the observable random variable say, $X_1, X_2, ...X_n$. The occurrence of instantaneous failures when some items are put on test giving $X_i = 0$ is quite common in electronic component and some other situations. Note that because of the limited accuracy of measuring failure time it is possible that we record $X_i = 0$ for some units although $P[X_i = 0 | \theta] = 0$. To accommodate such instantaneous failures, the model $\Im$ is modified to model $\mathcal{G}$ = {$G(x,\theta, \alpha), x \geq 0, \theta \in \Omega, 0 < \alpha < 1$}, where

$$G(x;\theta,\alpha) = \begin{cases} 1-\alpha, & x = 0 \\ 1-\alpha+\alpha F(x,\theta), & x > 0 \end{cases} \tag{3.2.1}$$

and $F(x,\theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp - \frac{1}{2\sigma^2}(y_i - \theta)^2 dy$ is df according to Normal distribution and α is the mixing proportion. The estimation of parameters in the above model is straight forward and depends on only the positive observations in the model. Thus

$$\hat{\alpha} = \frac{n-r}{n} \tag{3.2.2}$$

$$\hat{\theta} = \frac{\sum_{x_i>0} x_i}{n-r} \quad \text{and} \quad \hat{\sigma}_0^2 = \frac{\sum_{x_i>0}(x_i - \bar{x})^2}{n-r} \tag{3.2.3}$$

are easily obtainable. r denotes number of units that fail instantaneously. As we are considering life times of an object we get non-negative observations.

3.2.2 Normal with early failures

As we have already defined early failures in chapter 2, section (2.3), we can directly write the likelihood of this model as

$$L(x,\alpha,\theta)=\left[1-\alpha+\alpha F(\delta,\theta)\right]^{r}\left(\alpha\left[1-F(\delta,\theta)\right]\right)^{n-r}\prod_{x_i>\delta}\frac{f(x_i,\theta)}{1-F(\delta,\theta)} \tag{3.2.4}$$

where

$$F(\delta,\theta)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\delta}\exp-\frac{1}{2\sigma^2}(x_i-\theta)^2\,dx$$

that is, the likelihood of the sample under $g_1 \in \mathcal{G}_1$ is the product of the likelihoods of r (inliers) and the conditional likelihood of the sample given r which is same as the likelihood of (n-r) observations coming from the truncated version of $f \in \Im$ (or $g_1 \in \mathcal{G}_1$) restricted to (δ, ∞). Now r is binomial with probability of success given by $1-\alpha+\alpha F(\delta,\theta)$. For fixed θ and $\alpha \in [0,1]$ this binomial family is complete. Therefore, the optimal estimating equation for θ ignoring α is the conditional score function given r or $\frac{\partial \ln L_r}{\partial \theta}=0$, where $L_r=\frac{\prod_{x>\delta} f(x_i,\theta)}{1-F(\delta,\theta)}$. Hence optimal estimating equation for θ is given by equation (3.2.7). Thus, it is same as the estimator given by optimal estimating $\hat{\theta}$ equation for θ ignoring α. ML equations correspond to two parameter Normal models are given as

$$\ln L=r\ln\left[1-\alpha\bar{F}(\delta,\theta)\right]+(n-r)[\ln\alpha-\ln\sigma_1]-\frac{1}{2}\sum_{x_i>\delta}\frac{(x_i-\theta)^2}{\sigma_1{}^2} \tag{3.2.5}$$

$$\frac{\partial \ln L}{\partial \alpha}=0\Rightarrow\frac{-r\alpha\bar{F}(\delta,\theta,\sigma_1)}{1-\alpha\bar{F}(\delta,\theta,\sigma_1)}+\frac{(n-r)}{\alpha}=0 \tag{3.2.6}$$

$$\frac{\partial \ln L}{\partial \theta}=0\Rightarrow\frac{-r\alpha\frac{\partial}{\partial\theta}\bar{F}(\delta,\theta,\sigma_1)}{1-\alpha\bar{F}(\delta,\theta,\sigma_1)}+\sum_{r+1}^{n}\left(\frac{x_i-\theta}{\sigma_1^2}\right)=0 \tag{3.2.7}$$

and

$$\frac{\partial \ln L}{\partial \sigma_1} = 0 \Rightarrow \frac{-r\alpha \frac{\partial}{\partial \sigma_1} \bar{F}(\delta,\theta,\sigma_1)}{1-\alpha\bar{F}(\delta,\theta,\sigma_1)} - \left(\frac{n-r}{\sigma_1}\right) + \sum_{r+1}^{n} \frac{(x_i-\theta)^2}{\sigma_1^3} = 0 \qquad (3.2.8)$$

Here equations (3.2.7) and (3.2.8) may be solved simultaneously using Newton Raphson method. The above model gives reasonably good estimates of the parameters for δ fixed. See the example in the section (3.8), at the end of the chapter.

3.3 Normal with nearly instantaneous failures

With reference to equation **(**2.4.4) in chapter 2, normal with nearly instantaneous failures distribution can be written as

$$f(x) = p\delta_d(x - x_0) + q\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x-\theta}{\sigma_1}\right)^2\right) \quad ,p+q=1,\ 0<p<1 \qquad (3.3.1)$$

$$\sigma_1 > 0, \quad -\infty < \theta < +\infty$$

where

$$\delta_d(x-x_0) = \begin{cases} \frac{1}{d}, & x_0 \le x \le x_0 + d \\ 0, & otherwise \end{cases} \quad , \qquad (3.3.2)$$

for sufficiently small *d.* Here the mixing proportion $p>0$. Also note that

$$\delta(x - x_0) = \lim_{d\to 0} \delta_d(x - x_0) \qquad (3.3.3)$$

Since

$$f_1(x) = \delta_d(x - x_0)$$

and

$$f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x-\theta}{\sigma_1}\right)^2\right) \ , \ \sigma_1 > 0, \quad -\infty < \theta < +\infty$$

where $f(x)$ is given by

$$f(x) = p\,f_1(x) + q\,f_2(x) \quad \text{where} \quad p+q=1,\ 0<p<1. \tag{3.3.4}$$

and the corresponding survival function and hazard function of the mixture distribution are

$$R(x) = pR_1(x) + qR_2(x) \tag{3.3.5}$$

and

$$h(x) = \frac{pf_1(x)+qf_2(x)}{pR_1(x)+qR_2(x)} \tag{3.3.6}$$

respectively.

The components of $R(x)$ and $h(x)$ can be obtained as

$$R_1(x) = \begin{cases} 1, & 0 \le x < x_0 \\ \dfrac{d+x_0-x}{d}, & x_0 \le x \le x_0+d \\ 0, & x \ge x_0+d \end{cases} \tag{3.3.7}$$

and

$$R_2(x) = 1 - F_2(x) \qquad x > x_0 + d \tag{3.3.8}$$

$$h_1(x) = \begin{cases} 0, & 0 \le x < x_0 \\ \dfrac{1}{d+x_0-x}, & x_0 \le x \le x_0+d \\ \infty, & x \ge x_0+d \end{cases} \tag{3.3.9}$$

and

$$h_2(x) = \frac{\dfrac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-\theta}{\sigma_1}\right)^2\right)}{1-F_2(x)} \tag{3.3.10}$$

As a special case of the model, we obtain the Normal with "nearly instantaneous failure" model, when $t_0 = 0$ in equation (3.3.2). Accordingly the simplified expressions of the components in the failure rate and survival functions are

$$h_1(x) = \begin{cases} \dfrac{1}{d-x}, & 0 \le x \le d \\ \infty, & x > d \end{cases} \tag{3.3.11}$$

and its survival rate function in equation (3.3.7) is given as

$$R_1(x) = \begin{cases} \dfrac{d-x}{d}, & 0 \le x \le d \\ 0, & x > d \end{cases} \tag{3.3.12}$$

Thus the Normal model with "nearly instantaneous failure" occurring uniformly over [0, *d*] has survival function

$$R(x) = \begin{cases} \dfrac{p(d-x)}{d} + q\left[1 - F_2(x)\right], & 0 \le x \le d \\ q\left[1 - F_2(x)\right], & x > d \end{cases} \tag{3.3.13}$$

and

$$h(x) = \begin{cases} \dfrac{p}{p(d-x)+dq(1-F_2(x))}\left[1 - \dfrac{dp}{p(d-x)+dq(1-F_2(x))}\right]\dfrac{f_2(x)}{R_2(x)}, & 0 \le x \le d \\ \dfrac{qf_2(x)}{R_2(x)}, & x > d \end{cases} \tag{3.3.14}$$

Nearly instantaneous calculations are performed for the example in section (3.8).

3.3.1 Graphs

In various figures below we provide the graphs for $f(x)$, $R(x)$ and $h(x)$ for left values of mixing proportions and parametric values.

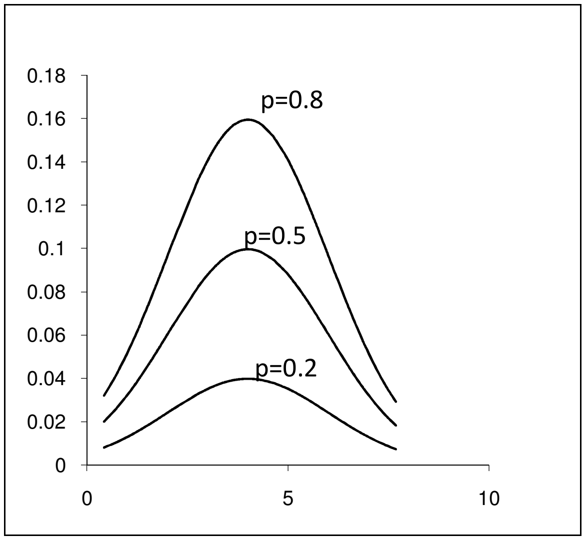


Fig. 3.3.1. Density function for μ = 4 and σ = 2

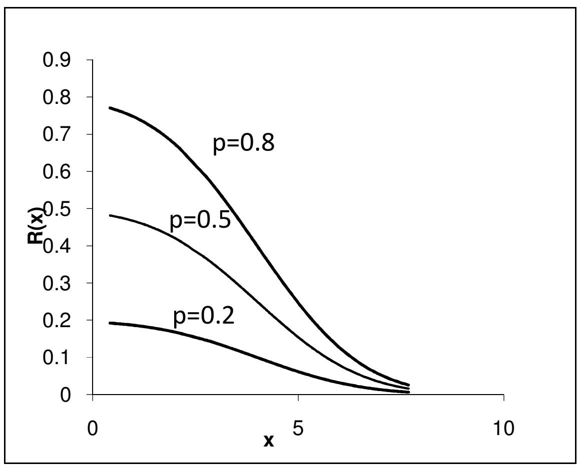


Fig. 3.3 2. Reliability function μ = 4 and σ = 2

Graph (3.3.4) and (3.3.5) are plotted on the basis of random sample generated from mixture of two normal distributions. From the graph (3.3.4) we can clearly identify two symmetrical curves, where first curve has inlier distribution with mean 4 remarkably less than second curve which can be considered as target distribution with mean 20. Graph (3.3.5) is known as normal quantile-quantile (Q-Q) plot. A sample from single normal distribution should produce a linear plot on this graph, which is not in our case. Hence both the graph clearly represents the presence of two groups.

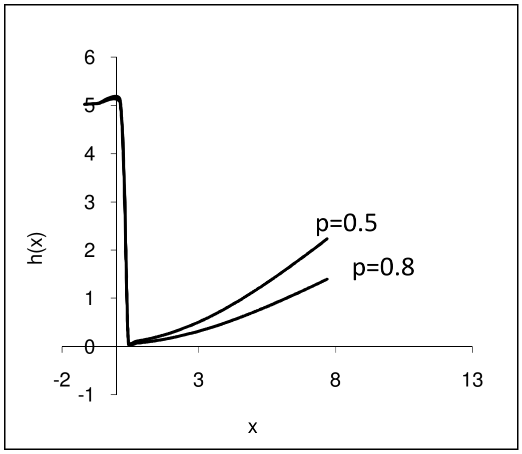


Fig. 3.3.3. Failure distribution for μ = 4 and σ = 2

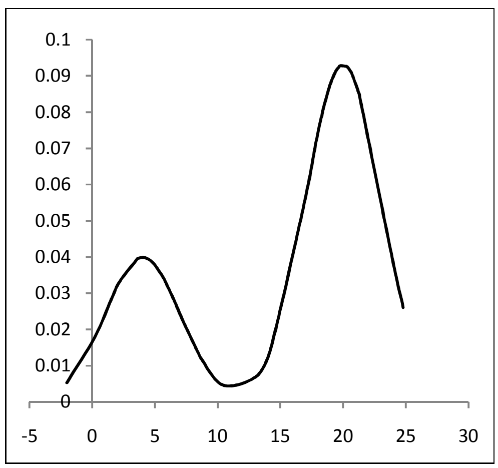


Fig 3.3.4. Density function of mixture of inliers and target distributions

3.4 Inlier detection methods

Here we obtain number of inliers for different data set by various methods, viz identified inlier model, labeled slippage methods and information criteria.

3.4.1 Identified inlier model (M_k)

Referring to equation (2.5.14) of section (2.5.2) from chapter 2 the identified inliers model with $g(x)$ as inliers and $f(x)$ as target distribution is written as

$$L(x|\phi,\theta,v,r)=\prod_{i=1}^{r}g(x_i)\prod_{i=r+1}^{n}f(x_i) \tag{3.4.1}$$

$$= \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi}\sigma_0} \exp - \frac{1}{2}\left(\frac{x_i - \phi}{\sigma_0}\right)^2 \prod_{i=r+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} \exp - \frac{1}{2}\left(\frac{x_i - \theta}{\sigma_1}\right)^2 \quad (3.4.2)$$

The likelihood function in (3.4.2) assumes that between the experiments when units are placed on test we do not know which of the units fail instantaneously. Equivalently $X_{i_1} = 0, X_{i_2} = 0, \ldots X_{i_r} = 0$ which fail early i.e. those units whose failure time distribution is $g\left(x_{(i)}, \phi\right)$ with failure rate much larger than that of the failure time distribution of the target population whose failure rate is considerably smaller. The identification is done as follows: evaluate for each fixed r where r = 0,1,2,...n-1 the maximum likelihood equation $\hat{L}_r$, and then consider $\hat{r}$ being that value of r for which likelihood is maximum. The computation for example of detection of inliers is done in section (3.5) and (3.8).

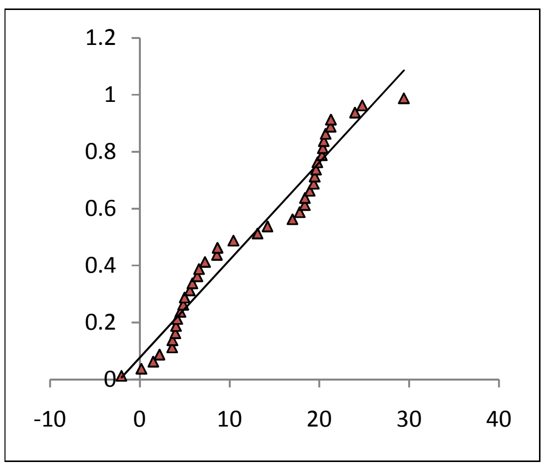


Fig. 3.3.5. Normal Probability Plot for mixture of two distributions

3.4.2 Inlier detection in Labeled slippage model (*L*$_k$)

With $g(x)$ and $f(x)$ as described in section (3.4.1), the likelihood under labeled slippage model referring to section (2.5) and substituting in equation (2.5.1), gives

$$\ln L = r_0 \ln(1-p) + (n-r_0)\ln p - \ln\varphi_{r_1}(\phi,\theta) + n\ln\sigma \; - \frac{\sum_{i=1}^{r_1}\left(x_{(i)}-\phi\right)^2}{2\sigma^2}$$
$$-\frac{\sum_{i=r_1+1}^{n}\left(x_{(i)}-\theta\right)^2}{2\sigma^2} \tag{3.4.3}$$

and the corresponding likelihood equations are

$$\frac{\partial \ln L}{\partial p} = \frac{-r_0}{(1-p)} + \frac{(n-r_0)}{p} = 0 \tag{3.4.4}$$

$$\frac{\partial \ln L}{\partial \phi} = -\frac{\partial}{\partial \phi}\ln\varphi_{r_1}(\phi,\theta) + \frac{\sum_{i=1}^{r_1} x_{(i)}}{r_1} \tag{3.4.5}$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{\partial}{\partial \theta}\ln\varphi_{r_1}(\phi,\theta) + \frac{\sum_{i=r_1+1}^{n} x_{(i)}}{(n-r_0-r_1)} \tag{3.4.6}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = 0 \Rightarrow \hat{\sigma} = \frac{\sum_{i=1}^{r_1}\left(x_{(i)}-\phi\right)^2 + \sum_{i=r_1+1}^{n}\left(x_{(i)}-\theta\right)^2}{n} \tag{3.4.7}$$

Here (3.4.4) can be solved to get the estimate of p as $\hat{p} = (n-r_0)/n$. The equations (3.4.5) and (3.4.6) contains gamma and digamma functions. The function

$$\varphi_{r_1}(\phi,\theta) = \frac{(n-r_0-r_1)}{\sqrt{2\pi}\sigma}\int_0^{\infty}\{G(x)\}^{r_1}\left[\bar{F}(x)\right]^{n-r_0-r_1} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}dx$$

where *G(x)* and *F(x)* are cumulative distribution functions of inlier and target population. The function $\varphi_{r_1}(\phi,\theta)$ is difficult to evaluate and can only be evaluated using some numerical method.

3.4.3 Information criterion for detection of inliers

As defined in chapter 2, section (2.6) here for Normal distribution, we have *SIC* for model with no inliers as

$$SIC(0) = 2n\log\sigma_1 + \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\sigma_1}\right)^2 + p\log n \tag{3.4.8}$$

and model with *r* inliers is defined as

$$SIC(r) = 2r\log\sigma_0 + 2(n-r)\log\sigma_1 + \sum_{i=1}^{r}\left(\frac{x_i - \phi}{\sigma_0}\right)^2 + \sum_{i=r+1}^{n}\left(\frac{x_i - \theta}{\sigma_1}\right)^2 + p\log n \tag{3.4.9}$$

The estimate of inliers say *r* is such that $SIC(r) = \min_{1 \le r \le n} SIC(r)$.

Here we use three information criteria such as *SIC, BIC* and *HQ* already defined in chapter 2. Hence *SIC* = $-2\ln L(\Theta)$+p ln n, *BIC* = $-\ln L(\Theta) + \frac{0.5p\ln(n)}{n}$ *and* *HQ* = $-\ln L(\Theta) +$ p $\ln[\ln(n)]$ can be used to detect the inliers, where $L(\Theta)$ the maximum likelihood function and *p* is the number of free parameters that need to be estimated under the model. We now illustrate this method using the simulated example discussed in the next section. Table (3.5.2) also presents the parameter estimates and the information criterion values.

3.5 Simulation study

To Illustrate the method of identifying inliers we have generated 15 independent random samples, where 5 of them are coming from normal distribution with parameter mean $\phi = 4$ and variance σ_0^2 = 2 and remaining ten observations from Normal distribution with parameter mean $\theta = 20$ and variance σ_1^2 = 2. The sample values are 1.44852, 3.667636, 3.949972, 5.548854, 6.017887, 17.61194, 19.26654, 20.09814, 20.23482, 20.36071, 20.64048, 21.08915, 21.26954, 22.53701 and 24.23439. We note that $SIC(0) = 58.4562 > SIC(5) = \min_{1 \le r \le n} SIC(r) = 34.85999$.

Table 3.5.1. The Likelihood and Information criterions

r	L	SIC	BIC	HQ
2	-38.1951	69.82944	-3.46217	-2.64648
3	-34.5019	62.44302	-3.36048	-2.54479
4	-31.2064	55.85195	-3.26009	-2.44439
5	**-20.7104**	**34.85999**	**-2.8501**	**-2.03441**
6	-26.054	45.54709	-3.07963	-2.26394
7	-28.546	50.53121	-3.17098	-2.35529
8	-30.997	55.43326	-3.25336	-2.43766
9	-33.0941	59.62746	-3.31882	-2.50313
10	-34.9391	63.31742	-3.37307	-2.55738
11	-36.6837	66.80655	-3.4218	-2.6061
12	-38.4748	70.38878	-3.46947	-2.65377
13	-39.6796	72.79842	-3.5003	-2.68461

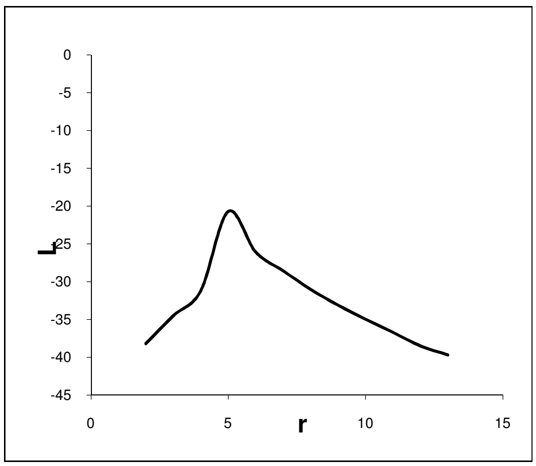


Fig. 3.5.1. Likelihood plot

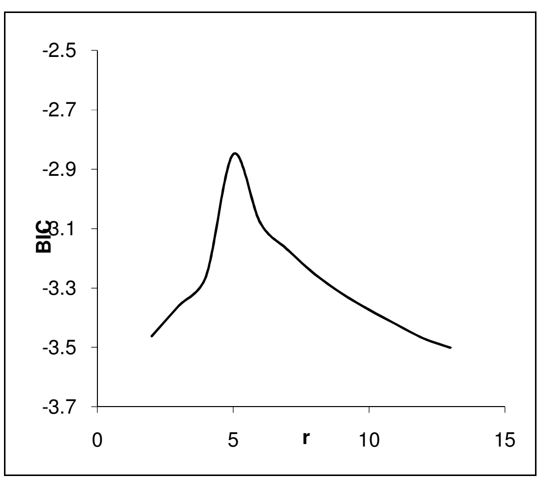


Fig. 3.5.2. *BIC* plot

A similar conclusion can be drawn in the case of other information criterions BIC and HQ also. Hence $r=5$ and the estimates are $\hat{\phi}=4.126574$, $\hat{\sigma}_0=1.803727$ $\hat{\theta}=20.73427$, and $\hat{\sigma}_1=1.783219$ respectively. The graphical representations of the likelihood and *BIC* plots are given in figure (3.5.1) and (3.5.2).

Next, we carried out an experiment with 1000 samples each of size 15 and number of inliers as 3, 4, 5 and 6 each with $\phi=3$ and $\theta=6, 9, 12$ and 15. The table (3.5.2) entitled power of SIC procedure presents the number of times the SIC procedure correctly identified the number of inliers in proportion to total number of samples. The values clearly indicate the effectiveness of the method in detecting the inliers. One of the important problem while detecting the inliers is the masking effect, where masking effect is defined as the loss of power due to wrong detection of more than one inliers.

Table 3.5.2. Power of SIC procedure

θ/ϕ \ r	2	3	4	5
3	0.570	0.720	0.700	0.550
4	0.460	0.480	0.490	0.440
5	0.460	0.460	0.460	0.462
6	0.410	0.420	0.430	0.410

3.6 Testing of hypothesis for inliers

After detection of number of inliers, it is necessary to test whether the methods used for detection are valid or not. Hence different tests are applied to test whether data truly represents our model of mixture of inliers and target population.

3.6.1 Sequential Probability Ratio Test (SPRT) to detect number of inliers

We want to test the hypothesis whether sample observations belong to inliers population from $N\left(\phi,\sigma_0^2\right)$ against hypothesis that it belongs to target population from $N\left(\theta,\sigma_1^2\right)$, assuming $\sigma=\sigma_0=\sigma_1$.

H_0 : sample observations are taken from normal population with mean ϕ

H_1 : sample observations are taken from normal population with mean θ

We use SPRT test given as follows:

The likelihood ratio λ_m is given by $\lambda_m = \frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln \lambda_m = \sum_{i=1}^{m} \ln \frac{f\left(x_{(i)}, \theta\right)}{g\left(x_{(i)}, \phi\right)}$$

$$= \frac{m\left(\phi^2 - \theta^2\right) + 2\left(\theta - \phi\right)\sum_{i=1}^{m} x_{(i)}}{2\sigma^2}, \quad m = 1,2,....n \tag{3.6.1}$$

For deciding number of inliers *r*, first arrange the observations in ascending order and then we continue to take likelihood ratio for m= 1, 2…. ,*n* by including observations one by one till we reject H_0. That is

If $\sum_{i=1}^{m} z_{(i)} \leq \ln B$ then accept H_0 and take the next observation.

and

If $\sum_{i=1}^{m} z_{(i)} \geq \ln A$ reject H_0 and stop.

The corresponding value of *m* represents the first observation from target population and number of inliers $\hat{r} = m - 1$. *A* and *B* are given as

$$B = \frac{\beta}{1-\alpha}, \quad A = \frac{1-\beta}{\alpha} \tag{3.6.2}$$

where α represents probability of type I error and β represents probability of type II error.

Test criteria for rejection of H_0 is

$$\ln \lambda_m > \ln A \Rightarrow \sum_{i=1}^{m} x_{(i)} > \frac{\sigma^2}{(\theta - \phi)} \ln A + \frac{m}{2}(\phi + \theta) \tag{3.6.3}$$

Corresponding value of m for which H_0 was accepted last becomes number of inliers r. The criteria is applied in example in section (3.8).

3.6.2 Modified likelihood ratio test

The study of the modified likelihood approach to finite normal mixture models with a common and unknown variance in the mixing components and a test of the hypothesis of a homogeneous model versus a mixture on two or more components was done by Chen and Kalbfleisch (2005). Here we use it to study the test for hypothesis

H_0 : sample observations are taken from single target normal population with mean θ

H_1 : sample observations are taken from mixture of inliers with mean ϕ and target distribution with mean θ.

We define $M_1 = \{F(x) : x \sim N(\theta, \sigma^2)\}$ i.e. all observations come from target population. $M_2 = \{F(x) = (1-p)F_1(x) + pF_2(x)\}$ *i.e.* X comes from mixture of two Normal distribution where $F_1(x)$ and $F_2(x)$ are distribution functions of inliers and target population, respectively, as defined in previous section.

Then the null hypothesis proceeds with testing $H_0 : p = 1$ against $H_1 : p < 1$ or in other words a test of the hypothesis $X \in M_1$ versus $X \in M_2$. The usual likelihood (LRT) statistics is given by

$$\ln\lambda = 2\left[\sup_{\theta, X \in M_1} \ln(\theta, X) - \sup_{\phi,\theta,X \in M_2} \ln(\phi,\theta,X)\right] \tag{3.6.4}$$

Due to non-regularity of the finite mixture models ln λ does not have usual chi-squared distribution. Therefore we proceed with a modified likelihood approach where the quantity $\ln(\phi,\theta,X)$ is replaced as

$$m\ln(\phi,\theta,X) = \ln(\phi,\theta,X) + c\ln\{4p(1-p)\} \tag{3.6.5}$$

where c is a positive constant. The purpose of the penalty term $c\ln\{4p(1-p)\}$ is to restore regularity to the problem by avoiding estimate of p on or near the boundary. Let $\ln(\hat{\theta},\hat{X}_1)$ maximizes $m\ln(\theta,X)$ for $X \in M_1$ and $\ln(\hat{\phi},\hat{\theta},\hat{X}_2)$ maximizes $m\ln(\phi,\theta,X)$ for $X \in M_2$. Thus modified likelihood ratio statistic is

$$\ln\hat{\lambda} = 2\left[\ln(\hat{\theta},\hat{X}_1) - \ln(\hat{\phi},\hat{\theta},\hat{X}_2)\right] \tag{3.6.6}$$

The null hypothesis is rejected for values of $\ln\hat{\lambda}$ that are sufficiently large. Here $\ln\hat{\lambda}$ follows $\chi^2_{(2)}$ distribution.

3.6.3 Most powerful test for detection of inliers

The most powerful test for testing the hypothesis as given in (1.6.1) whether the sample is from single population, we frame the hypothesis with common parameter μ

$H_0 : \mu = \phi$ i.e sample observations are from inliers normal population

$H_1 : \mu = \theta$ i.e sample observations are from target normal population

where μ is the mean of normal population and $\theta > \phi$.

Then the most powerful test is as given below

$$\psi(x) = \begin{cases} 1, & \dfrac{P_1(x)}{P_0(x)} > C\alpha \\ 0, & \dfrac{P_1(x)}{P_0(x)} < C\alpha \end{cases} \tag{3.6.7}$$

which can be simplified as

$$\psi(x) = \begin{cases} 1, & \sum_{i=1}^{n} x_i > \dfrac{C\alpha\sigma^2}{(\theta-\phi)} + \dfrac{n(\theta+\phi)}{2} \\ 0, & o.w \end{cases} \tag{3.6.8}$$

where C_α is such that the test attains level of the test when H_0 is true. Thus we reject H_0 for large values of the $\sum_{i=1}^{n} x_i$ with $C_\alpha = \phi + \sigma z_\alpha$.

3.6.4 F- test to test whether data contains inlier observations

To test whether the data is taken from single normal population or from mixture of inlier and target (both normal) distributions, we proceed with the F-test as follows

$H_0 : x_1, x_2, ...x_n$ are independent and follows $N\left(\theta, \sigma^2\right)$
$H_1 : x_{(1)}, x_{(2)}, ...x_{(r)}$ follows $N\left(\phi, \sigma_0^2\right)$ and $x_{(r+1)}, x_{(r+2)}, ...x_{(n)}$ follows $N\left(\theta, \sigma_1^2\right)$
where $\phi < \theta$.

Then test statistic obtained by Titterrington(1985) gives the maximum ratio of between sum of squares to within sum of squares as

$$F_{\max} = \frac{\max n_1 n_2 \left(\bar{x}_1 - \bar{x}_2\right)^2}{\left[\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2\right]\left(n_1 + n_2\right)} \tag{3.6.9}$$

where the maximum is over all partitioning of data set into two groups..

For detection of inliers, we find F_{max} for all possible values of r =1, 2,...n-1. The number of inliers r will be detected for which corresponding value of F_{max} is maximum.

3.7 Masking effect on tests for inliers

Let $X_1, X_2....X_n$ be sequence of n independent random variables with some known FTD. Under the null hypothesis H_0 these random variables are identically distributed with df F whereas under alternative hypothesis H_1, discordant observations (inliers) arise from population df *G*. The df of *G* is assumed to be of same form as that of *F* with a change in location or scale parameter by an unknown quantity λ. This parameter is called discordancy parameter, measuring the degree of discordancy. Under H_1 it is assumed that one of the observation follows df *G*. Let T(x) be a test statistics to detect a single discordant observation with critical region *A(n,α).* Due to lack of information about the number of discordant observations present in the sample, however, the true situation may not be specified by H_1 and more than one discordant observation may be present in the sample. In such cases test statistics T(x) suggested for detection of a single discordant, may fail to detect a single inlier as discordant even when additional discordant observations are present in the sample. Such a phenomenon is called masking effect.

All tests for detecting a single inlier, H_0 against H_1 are based on symmetric functions of observations or on functions of order statistics. In the k-inlier model, the joint distribution of order statistics $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ is same as that under the exchangeable model introduced by Kale (1975) where it is assumed that any set $X_{i_1}, X_{i_2}, \ldots X_{i_k}$ has priori equal probability of being independent and identically distributed as G_λ and the remaining (*n*-*k*) observations are distributed as *F*, the distribution function of target population.

In exchangeable model $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ has minimum posterior probability of coming from G_λ such that $\frac{\partial G_\lambda}{\partial F}$ is the decreasing function in X. The limiting masking effect by Bendre and Kale (1985) can be studied by assuming $X_{(1)}, X_{(2)}, \ldots X_{(k)}$ correspond to observation coming from $N\left(\mu - \lambda\sigma, \sigma^2\right)$ and then taking limit as $\lambda \to \infty$.

$$h\left(x_{(1)}, x_{(2)} \ldots x_{(n)}\right) = \frac{k!(n-k)!}{\varphi_\lambda(1,2,3\ldots\ldots k)} \prod_{i=1}^{k} g_\lambda(x_i) \prod_{i=k+1}^{n} f(x_i) \;, \qquad (3.7.1)$$

$$-\infty < x_{(1)} < x_{(2)} \ldots < x_{(n)} < \infty$$

Also f and g_λ are probability density functions of $N\left(\mu,\sigma^2\right)$ and $N\left(\mu-\lambda\sigma,\sigma^2\right)$ respectively. Thus masking effect on any test statistics *T(x)* with critical region *A*(*n,α*), for Labelled slippage model L_{sk} for *k ≥ 1*, is obtained as

$$\lim_{\lambda\to\infty} P\left[T(x)\in A(n,\alpha)/L_{sk}\right]=\lim_{\lambda\to\infty}\int_{A(n,\alpha)} h\left(x_{(1),}x_{(2),\ldots\ldots}x_{(n)}\right)dx_{(1)}\ldots\ldots dx_{(n)} \qquad (3.7.2)$$

Under L_{sk} as $\lambda\to\infty$, $X_{(n-k+1)},X_{(n-k+2)},\ldots..X_{(n)}$ behave as order statistics of a sample of size (*n*-*k*) from $N\left(\mu,\sigma^2\right)$ and $X_{(1)},X_{(2)},\ldots..X_{(k)}$ diverge to zero. However if $T\left(x_{(1)},x_{(2)},\ldots..x_{(k)}\right)$ is a function whose distribution does not depend on λ then T converges in distribution to a proper random variable as $\lambda\to\infty$.

3.7.1. Limiting masking effect

For single inlier in left tail, that is to test whether $x_{(1)}$ is an inliers, Grubbs proposed a test proposed by Bendre and Kale (1987).

$$G=\frac{\sum_{i=2}^{n}\left(x_{(i)}-\bar{x}_n\right)^2}{\sum_{i=1}^{n}\left(x_{(i)}-\bar{x}\right)^2}, \qquad (3.7.3)$$

where $\bar{x}_n=\dfrac{\sum_{i=2}^{n}x_{(i)}}{n-1}$ and $\bar{x}=\dfrac{\sum_{i=1}^{n}x_{(i)}}{n}$

The maximum studentized residual T is given by

$$T=\frac{\dfrac{(n-1)}{n}}{\left[\sum\dfrac{\left(x_{(i)}-\bar{x}_{(n)}\right)^2}{\left(x_{(n)}-\bar{x}_{(1)}\right)^2}+\dfrac{(n-1)}{n}\right]^{\frac{1}{2}}} \qquad (3.7.4)$$

where the sum is over $i = 2,3......n$. Since under L_{s1} corresponds to the one inlier observation coming from $N\left(\mu-\lambda\sigma,\sigma^2\right)$ and $\frac{\left(x_{(i)}-\overline{x}_{(n)}\right)^2}{\left(x_{(1)}-\overline{x}_{(n)}\right)^2} \to 0$ in probability as $\lambda \to \infty$ for i =2,3,4........n and therefore $T \to \left[\frac{n-1}{n}\right]^{\frac{1}{2}}$ in probability as $\lambda \to \infty$. Hence as $\lambda \to \infty$, $\lim P_1^G(\lambda)=1$ where $P_1^G(\lambda)$ is the power function of Grubb's test. To study $\lim P_2^G(\lambda)=\lim P\left[T<t_{n,\alpha} \mid L_{sk}\right]$ as $\lambda \to \infty$ we write

$$T = \frac{Y_{(1)}-\frac{k}{n}}{\left[\sum Y_{(1)}^2 - 2k\frac{\sum Y_{(i)}}{n}+\frac{k^2}{n}\right]^{\frac{1}{2}}} \tag{3.7.5}$$

sums are over i =1, 2,...n and where

$$Y_{(i)} = \frac{\left(x_{(i)}-\overline{x}_{(k)}\right)}{\left(\overline{x}'_{(n-k+1)}-\overline{x}_{(k)}\right)} \qquad i=1,2,......n \tag{3.7.6}$$

with $\overline{x}'_{(n-k+1)}$ is the mean of $x_{(k+1)},x_{(k+2)},.....x_{(n)}$ and $\overline{x}_k$ is the mean of $x_{(1)},x_{(2)},.....x_{(k)}$ Therefore $Y_{(i)} \to 0$ in probability for i =1, 2.....k because the numerator of $Y_{(i)}$ is a proper r.v., while denominator diverges to infinity. For i =1, 2,......k, we observe that $Y_{(i)}-1=\frac{\left(x_{(i)}-\overline{x}'_{(n-k+1)}\right)}{\left(\overline{x}_{(n-k+1)}-\overline{x}_{(k)}\right)}$ is such that the numerator has a distribution independent of λ and therefore converges to a proper random variable, but denominator diverges to infinity and hence $Y_{(i)} \to 1$ in probability as $\lambda \to \infty$. Therefore under L_{sk} as $\lambda \to \infty$,

$$T \to \left[\frac{(n-k)}{nk}\right]^{\frac{1}{2}}$$

and

$$\lim P_2^G(\lambda) = \begin{cases} 1, & \left[\frac{(n-k)}{nk}\right]^{\frac{1}{2}} < t_{n,\alpha} \\ 0 & o.w. \end{cases} \tag{3.7.7}$$

Thus Grubb's test is free from the limiting masking effect for $\left[\frac{(n-k)}{nk}\right]^{\frac{1}{2}} \geq t_{n,\alpha}$ and the performance of the test depends on the sample size *n* and the number of inliers. In general $t_{n,\alpha}$ is a decreasing function of the sample size and hence for large *n* with moderate *k* the test is free from the limiting masking effect. Table (3.7.1), presents the maximum number of inliers in a sample of size *n* upto which Grubb's test is free from the limiting masking effect.

Table. 3.7.1 Maximum inliers accommodated by Grubb's test

α	n =10	n = 15	n = 20	n = 25
0.01	1	1	1	2
0.05	1	2	2	2
0.10	1	2	2	3

3.8 Illustrations

3.8.1 Vannman's data

This example is based on a wood drying experiment. The data of Schedule 1 and 2 of Experiment 3 conducted by Vannman (1991). In both the case *n*=37. For data refer appendix.

Table (3.8.1) presents the estimates of the parameters of target distribution under instantaneous failure, early failures and nearly instantaneous models.

Table 3.8.1 Estimation for instantaneous failure, early failures and nearly instantaneous failures

Schedule		Instantaneous	Early failures	Nearly instantaneous
1 δ **=1.5**	$\hat{\theta}$	4.867917	7.352	5.076087
	$\hat{\sigma}_1$	4.398309	3.745867	4.374601
2 δ **=0.9**	$\hat{\theta}$	2.439	3.919167	3.0425
	$\hat{\sigma}_1$	2.606334	2.390099	2.581076

3.8.2 Rainfall data

The data, collected by Amutha and Porchelvan (2009), represents average monthly rainfall (in mm) during year 2004 and 2006 for the estimation of surface runoff in Malattar Sub-watershed which is a major tributary of Palar river. The watershed experiences tropical monsoon climate with normal temperature, humidity and evaporation throughout the year. The data was published in Journal of the Indian Society of Remote Sensing. For our illustration's purpose we reproduce two sets of data from the above paper.

Set 1 (2004) : 3.40, 0.00, 0.00, 15.80, 232.80, 8.80, 123.20, 47.00, 154.00, 103.20, 89.80 and 12.20.

Set 2 (2006) : 0.00, 0.00, 21.40, 60.20, 53.86, 93.20, 27.80, 45.40, 205.40, 101.20, 128.20 and 0.00.

We have combined the two sets together and arranged in ascending order to obtain inlier detection discussed in section (3.3), (3.4) and (3.6). Table (3.8.2), represents the value of inlier numbers r, likelihood, *SIC(r), BIC(r) HQ(r)* and modified test statistics for different values of *r*.

Table 3.8.2. Detection of number of inliers

r	Likelihood	SIC	BIC	HQ	$\ln\hat{\lambda}$
2	-39.796	85.9489	-3.5514	-2.5275	7.58094
3	-36.174	78.7048	-3.4559	-2.4321	14.8250
4	-32.756	71.8689	-3.3567	-2.3328	21.6609
5	-30.897	68.1503	-3.2982	-2.2744	25.3795
6	-28.634	63.6245	-3.2222	-2.1983	29.9053
7	-27.532	61.4194	-3.1829	-2.1591	32.1104
8	-25.643	57.6421	-3.1119	-2.0880	35.8877
9	**-23.759**	**53.8748**	**-3.0356**	**-2.0117**	**39.6550**
10	-27.474	61.3047	-3.1808	-2.1570	32.2251
11	-28.165	62.6857	-3.2057	-2.1818	30.8441
12	-29.31	64.9769	-3.2455	-2.2217	28.5529
13	-29.606	65.5676	-3.2555	-2.2317	27.9622
14	-30.516	67.3886	-3.2858	-2.2620	26.1412
15	-31.102	68.5595	-3.3048	-2.2810	24.9702
16	-32.072	70.5005	-3.3356	-2.3117	23.0293
17	-33.225	72.8055	-3.3709	-2.3470	20.7243
18	-35.026	76.4082	-3.4237	-2.3998	17.1216
19	-36.531	79.4180	-3.4657	-2.4419	14.1118
20	-37.807	81.9707	-3.5001	-2.4762	11.5591
21	-39.347	85.0499	-3.5400	-2.5161	8.47988
22	-40.865	88.0857	-3.5778	-2.5540	5.44412

Table (3.8.2) gives us *SIC(0)* = 99.45467 > *SIC(9)* = min *SIC(r)* = -23.759. The likelihood is maximum for *r = 9*. The corresponding estimates of the parameter are $\hat{\phi}$ = 0.72778, σ_0 = 0.45686 and $\hat{\theta}$ = 7.352 , σ_1 = 3.74587. For modified likelihood ratio test also maximum $\ln\hat{\lambda}$ corresponds to *r = 9*. We observe that, *SIC(0)* = 183.2181 > *SIC(6)* = min *SIC(r)* = 173.5757 . Also the likelihood is maximum for *r = 6*. The corresponding estimates of the parameter are $\hat{\phi}$ = 14.9, σ_0 = 8.78886 and $\hat{\theta}$ = 111.182, σ_1 = 58.0748. For modified likelihood ratio test also maximum $\ln\hat{\lambda}$ corresponds to *r = 6*. For SPRT, we test H_0: $\phi = 15$ against H_0: $\phi > 15$. For which we considered (α, β) =(0.02, 0.05). Then *ln A* = -2.5647 and *ln B* = -2.9755, and the computed statistics value is $\frac{\sigma^2}{(\theta-\phi)}\ln A + \frac{m}{2}(\phi+\theta)$ = 101.2454 .

Table 3.8.3 Estimates of parameters and detection of r

r	Likelihood	*SIC*	*BIC*	*HQ*	$\ln\hat{\lambda}$	$\sum_{i=1}^{m} x_{(i)}$
2	-91.4964	188.8817	91.57392	91.49643	-5.663600	12.200
3	-88.8329	183.5547	88.91039	88.83290	-0.336540	24.400
4	-86.5127	178.9142	86.59016	86.51268	4.303914	40.200
5	-84.9457	175.7802	85.02315	84.94566	7.437946	61.600
6	**-83.8434**	**173.5757**	**83.92090**	**83.84341**	**9.642444**	89.400
7	-85.0201	175.9291	85.09758	85.02010	7.289072	134.80
8	-84.4475	174.7838	84.52495	84.44746	8.434343	181.80
9	-83.9214	173.7317	83.99890	83.92141	9.486446	235.66
10	-84.2590	174.4069	84.33647	84.25899	8.811291	303.66
11	-85.9395	177.7679	86.01701	85.93953	5.450209	393.56
12	-86.7336	179.3560	86.81104	86.73355	3.862165	486.76
13	-87.4507	180.7903	87.52822	87.45073	2.427797	587.96
14	-87.6541	181.1971	87.73158	87.65410	2.021071	691.16
15	-88.6952	183.2794	88.77273	88.69525	-0.061230	814.36
16	-89.1081	184.1052	89.18562	89.10814	-0.887010	942.56
17	-89.4825	184.8538	89.55995	89.48247	-1.635670	1096.5

Hence we reject H_0 for first time when inlier r is 7 and conclude that number of inliers in the above data set, see table (3.8.3) is $\hat{r} = 6$.

Chapter 4

Inliers estimation in Weibull models

4.1 Introduction

One of the most important and widely used distributions to study lifetime of any component is Weibull distribution. The Weibull distribution is appropriate to describe the variation in the lifetimes of many different types of components. It has been used as model of lifetimes with diverse types of items such as Vacuum tubes, ball bearings and electrical insulation. In survival analysis, the Weibull distribution is better suited than the Gaussian distribution, because, it is defined only for positive time (engines fail after assembly), the mathematical operations particular to reliability theory are simpler (e.g. the function is easy to integrate analytically) and the shape of the function is more flexible, it ranges from a close Gaussian resemblance, to a skewed Gaussian, to a pure exponential distribution. It is also widely used in biomedical applications for e.g. in studies on time to occurrence of tumors in human population or in laboratory animals etc. It includes exponential distribution as special case. Also exponential distribution has been widely used as model in areas ranging from studies on the lifetimes of manufactured items to research involving survival or remission times in chronic diseases. In all the above examples we can get inliers and target observations as discussed in chapter 1.

As discussed in the previous chapters, here also our objective is to study the inliers and their detection procedure in Weibull distributions. This chapter deals with competition between two Weibull mixture models representing inliers and the target distribution.

If we denote $\underline{X}=(X_1,X_2....X_n)$ as realizations of a life test, then $\underline{X}=(X_I \cup X_T)$ where X_I is set of inlier observations (instantaneous and early failures) and X_T is set of observations coming from a target population. Since failure pattern of this situations usually discard the assumption of unimodal distribution, the usual method of modeling and inference procedures may not be accurate in practice. The prior objective in such situations is to decide how many inliers are present in the underlying model, and then study their inferences.

This article is organized as follows: In section (4.2) and (4.3), discussion of the UMVUE and identified inliers model, assuming both inlier distribution and target distribution as Weibull distribution is considered. The inference procedures when some of the parameters are known and unknown are considered. Section (4.3.3) deals with the inlier detection for labeled slippage model. The detection using information criterion, goodness of fit and data analysis are given in the subsequent sections.

4.2 Uniformly minimum variance unbiased estimator (UMVUE)

The UMVUE of mixture density of instantaneous and positive observation taken from Weibull distribution is obtained in this section. Based on above families a new family of df $\Im=\{F(x;\theta,p): x\geq 0,\theta\in\Omega,0<p<1\}$ is defined, such that

$$f(x;\theta,p)=\begin{cases}1-p+pf(x;\theta), & x=0\\ pf(x;\theta), & x>0\end{cases}$$

Hence the pdf of mixture family of instantaneous and Weibull distribution is obtained as

$$f(x;\theta,p)=(1-p)^{\delta}\,p^{(1-\delta)}\left[\beta\theta x^{\beta-1}\,e^{-\theta x^{\beta}}\right]^{(1-\delta)}$$

$$=\frac{\left(\beta x^{\beta-1}\right)^{(1-\delta)}\left[\exp(-\theta)\right]^{(1-\delta)x^{\beta}}\left[\frac{(1-p)}{p\theta}\right]^{\delta}}{\left[\frac{1}{p\theta}\right]} \tag{4.2.1}$$

which is a member of exponential family with $a(x)=\beta x^{\beta-1}$, $h(\theta)=\exp(-\theta)$, $g(\theta)=\frac{1}{\theta}$ and $d(x)=x^{\beta}$. We have $z=\sum_{x>0}(1-\delta)x^{\beta}$ and $n-r=\sum_{x>0}\delta_j$ which are jointly complete sufficient statistics for (θ, p). Since x^{β} has exponential distribution with parameter θ. The UMVUE of mixture density given by Singh (2007) is defined as

$$\varphi_x(z,r,n)=\begin{cases}\frac{B(z,r,n-1)}{B(z,r,n)}=\frac{n-r}{n}, & x=0,\ r=0,1,2....n-1\\ a(x)\frac{B(z-d(x),r-1,n-1)}{B(z,r,n)}, & x>0,\ z>d(x),\ r=1,2....n\end{cases}$$

which simplifies in Weibull as

$$\varphi_x(z,r,n)=\begin{cases}\frac{n-r}{n}, & x=0\\ \frac{r(r-1)\beta x^{\beta-1}}{nz}\left[1-\frac{x^{\beta}}{z}\right]^{r-2}, & 0<x^{\beta}<z,\ n>1\end{cases} \tag{4.2.2}$$

If $\beta=1$ in equation (4.2.2) one gets UMVUE of mixture density of instantaneous and positive observation from exponential distribution as

$$\varphi_x(z,r,n)=\begin{cases}\frac{n-r}{n}, & x=0\\ \frac{r(r-1)}{nz}\left[1-\frac{x}{z}\right]^{r-2}, & 0<x<z,\ n>1\end{cases} \tag{4.2.3}$$

4.3 Weibull identified inliers model

Weibull distribution is used medical studies dealing with fatal diseases, where one is interested in the survival time of individual with the disease, measured either from the date of diagnosis or some other starting point. It is possible that patient dies without getting treatment or has smaller survival rate than target group which has on average longer survival rate. The inlier detection is done for the following two models:

Model-1: Shape parameter β is same for both inliers and target distribution.

Model-2: Scale parameter θ is same for both inliers and target distribution.

4.3.1 Inlier detection when the shape parameter is identical

If we take the distribution function of inliers as

$$G(x) = 1 - \exp\left(-\phi x^{\beta}\right), \qquad x > 0,\ \phi > 0,\ \beta > 0. \tag{4.3.1}$$

and the distribution of target population is

$$F(x) = 1 - \exp\left(-\theta x^{\beta}\right), \qquad x > 0, \theta > 0, \beta > 0. \tag{4.3.2}$$

Then the likelihood of model can be written as

$$L(x \mid \phi, \theta, \beta) = \prod_{i=1}^{r} \beta\phi x^{\beta-1} e^{-\phi x^{\beta}} \prod_{i=r+1}^{n} \beta\theta x^{\beta-1} e^{-\theta x^{\beta}}$$

The estimates of the parameters are found by solving the following likelihood equations:

$$\frac{\partial \ln L}{\partial \phi} = 0 \Rightarrow \frac{r}{\phi} - \sum_{i=1}^{r} x_i^{\beta} = 0$$

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{n-r}{\theta} - \sum_{i=r+1}^{n} x_i^{\beta} = 0$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \phi \sum_{i=1}^{r} x_i^{\beta} \ln x_i + \theta \sum_{i=r+1}^{n} x_i^{\beta} \ln x_i - \frac{n}{\beta} - \sum_{i=1}^{n} \ln x_i = 0$$

As discussed earlier, the instantaneous failures are already identified and hence the proportion of such observations is not considered in the model. Using Newton Raphson method the estimates of ϕ, θ and β can be found.

4.3.2 Inlier detection when the scale parameters are identical

Here we consider the detection of inliers when shape parameters of both inlier and target distribution as same. The failure distribution for inliers is assumed to be

$$G(x) = 1 - \exp\left(-x^{\beta}\theta\right), \qquad x > 0,\ \theta > 0,\ \beta > 0. \tag{4.3.3}$$

and the distribution of target population is

$$F(x) = 1 - \exp\left(-x\theta\right), \qquad x > 0\ ,\theta > 0. \tag{4.3.4}$$

The likelihood estimates in this case are the solutions of

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{n}{\theta} - \sum_{i=1}^{r} x_i^{\beta} + \sum_{i=r+1}^{n} x_i = 0$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \sum_{i=1}^{r} x_i^{\beta} \ln x_i - \frac{r}{\beta} - \sum_{i=1}^{r} \ln x_i = 0$$

Since all the likelihood equations are non linear, they may be solved using Newton Raphson method, to get estimates of θ and β.

4.3.3 Labeled slippage inliers model for Model-1

With $g(x)$ and $f(x)$ as described above, the likelihood under labeled slippage model referring to section (2.5) and substituting in equation (2.5.1), gives

$$\ln L = r_0 \ln(1-p) + (n-r_0)\ln p - \ln\varphi_{r_1}(\phi,\theta) + r_1 \ln\phi + (\beta-1)\sum \log x_{(i)} - \phi\sum_{i=1}^{r_1} x_{(i)}^{\beta}$$
$$+(n-r_0-r_1)\ln\theta - \theta\sum_{i=r_1+1}^{n} x_{(i)}^{\beta}$$

and the corresponding likelihood equations are

$$\frac{\partial \ln L}{\partial p} = \frac{-r_0}{(1-p)} + \frac{(n-r_0)}{p} = 0 \tag{4.3.5}$$

$$\frac{\partial \ln L}{\partial \phi} = -\frac{\partial}{\partial \phi}\ln\varphi_{r_1}(\phi,\theta) + \frac{r_1}{\phi} - \sum_{i=1}^{r_1} x_{(i)}^{\beta} \tag{4.3.6}$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{\partial}{\partial \theta}\ln\varphi_{r_1}(\phi,\theta) + \frac{n-r_0-r_1}{\theta} - \sum_{i=r_1+1}^{n} x_{(i)}^{\beta} \tag{4.3.7}$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \phi\sum_{i=1}^{r} x_i^{\beta}\ln x_i + \theta\sum_{i=r+1}^{n} x_i^{\beta}\ln x_i - \frac{n}{\beta} - \sum_{i=1}^{n}\ln x_i = 0 \tag{4.3.8}$$

Here (4.3.5) can be solved to get the estimate of p as $\hat{p} = (n-r_0)/n$. The equations (4.3.6) to (4.3.8) contains gamma and digamma functions. Solving (4.3.6) and (4.3.7) simultaneously we get the estimate of ϕ and θ. The parameter p is orthogonal to $(\phi,\theta)'$. Now

$$\varphi_{r_1}(\phi,\theta) = (n-r_0-r_1)\int_0^{\infty}\left\{1-e^{-x^{\beta}\phi}\right\}^{r_1}\left[e^{-x^{\beta}\theta}\right]^{n-r_0-r_1}\beta\theta x^{\beta-1}dx$$

$$= \frac{(n-r_0-r_1)\theta}{\phi}\beta\left(r_1+1, \frac{(n-r_0-r_1)\theta}{\phi}\right)$$

$$= \left[\frac{(n-r_0-r_1)\theta}{\phi}\right]\left[\frac{\Gamma(r_1+1)\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}\right)}{\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}+r_1+1\right)}\right]$$

and

$$\ln\varphi_{r_1}(\phi,\theta)=C+\ln\theta-\ln\phi+\ln\Gamma(z)-\ln\Gamma((z+r_1+1))$$

where

$$z=([n-r_0-r_1]\theta)/\phi.$$

hence

$$\frac{\partial}{\partial\phi}\ln\varphi_{r_1}(\phi,\theta)=-\frac{1}{\phi}+\frac{\partial}{\partial\phi}\ln\Gamma(z)\frac{\partial z}{\partial\phi}-\frac{\partial}{\partial\phi}\ln\Gamma(z+r_1+1)\frac{\partial z}{\partial\phi}$$

$$=-\frac{1}{\phi}+\left[\psi(z)-\psi(z+r_1+1)\right]\left(-\frac{(n-r_0-r_1)\theta}{\phi^2}\right)$$

and

$$\frac{\partial}{\partial\theta}\ln\varphi_{r_1}(\phi,\theta)=-\frac{1}{\theta}+\frac{\partial}{\partial\theta}\ln\Gamma(z)\frac{\partial z}{\partial\theta}-\frac{\partial}{\partial\theta}\ln\Gamma(z+r_1+1)\frac{\partial z}{\partial\theta}$$
$$=-\frac{1}{\theta}+\left[\psi(z)-\psi(z+r_1+1)\right]\left(\frac{(n-r_0-r_1)}{\phi}\right)$$

where

$$\psi(z)=\frac{\partial}{\partial\phi}\ln\Gamma(z)\quad\text{and}\quad\Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx.$$

The result from Abramovitz and Stegun (1965) is

$$\left[\psi(z)-\psi(z+r_1+1)\right]=-\sum_{j=1}^{r_1}\frac{1}{z+j}. \qquad (4.3.9)$$

Using the above results, one obtain the likelihood equations as

$$\frac{\partial\ln L}{\partial\phi}=\frac{r_1+1}{\phi}-\frac{\theta(n-r_0-r_1)}{\phi^2}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=1}^{r_1}x_i^\beta=0 \qquad (4.3.10)$$

and

$$\frac{\partial\ln L}{\partial\theta}=\frac{(n-r_0-r_1-1)}{\theta}+\frac{(n-r_0-r_1)}{\phi}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=r_1+1}^{n}x_i^\beta=0 \qquad (4.3.11)$$

Now (4.3.8), (4.3.10) and (4.3.11) can be solved using Newton-Raphson method to get the estimates of ϕ, θ and β.

4.4 Inliers detection using information criterion

Here three information criteria are used to detect inliers, which are already discussed in chapters 2, section (2.5) such as Schawarz's Information criterion ($SIC = -2\ln L(\Theta) + p\ln n$), the Schawarz's Bayesian Information criterion ($BIC = -\ln L(\Theta) + 0.5\,\frac{p\, ln\, n}{n}$) and the Hannan-Quinn criterion defined as ($HQ = -lnL(\Theta)\ +\ p\ ln\left[ln(n)\right]$). Here $L(\Theta)$ the maximum likelihood function and p is the number of free parameters that need to be estimated under the model. Below we develop the procedure for SIC scheme. The following model of no inliers for Model-1 is given by

$$SIC(0) = -2n\ln\theta - 2n\ln\beta + 2\theta\sum_{i=1}^{n} x_{(i)}^{\beta} - 2(\beta-1)\sum_{1}^{n}\ln x_{(i)} + 2\ln n \qquad (4.4.1)$$

and the corresponding model with r inliers is

$$SIC(r) = -2r\ln\phi - 2(n-r)\ln\theta - 2n\ln\beta + 2\phi\sum_{i=1}^{r} x_{(i)}^{\beta} + 2\theta\sum_{i=r+1}^{n} x_{(i)}^{\beta} - 2(\beta-1)\sum_{1}^{n}\ln x_{(i)} + 3\ln n \quad (4.4.2)$$

Similarly, for Model-2, the model with no inliers is

$$SIC(0) = -2n\ln\theta + 2\theta\sum_{1}^{n} x_{(i)} + \ln n \qquad (4.4.3)$$

and corresponding model with r inliers is

$$SIC(r) = -2n\ln\theta - 2r\ln\beta - 2(\beta-1)\sum_{1}^{r}\ln x_{(i)} + 2\theta\sum_{1}^{r} x_{(i)}^{\beta} + 2\theta\sum_{r+1}^{n} x_{(i)} + 2\ln n \qquad (4.4.4)$$

The estimate of inliers say r is such that $SIC(r) = \min_{1\le r\le n} SIC(r)$, where $r, 1 \le r \le n-1$, is the unknown index of the inliers. According to the procedure, the Model with no inlier is selected if $SIC(0) < \min_{1 \le r \le n-1} SIC(r)$. And the Model with *r* inlier is selected if $SIC(0) > \min_{1 \le r \le n-1} SIC(r)$.

4.4.1 Simulation study

To illustrate the method of identifying inliers model the random samples of size 15 have been generated from Weibull distribution. The data under two models are as follows:

Model-1: Five observations are generated from Weibull with parameter $\phi = 0.50$ and $\beta = 1.1$ and remaining ten observations from Weibull distribution with parameter $\theta = 0.25$ and $\beta = 1.1$. The ordered observations are 0.1475, 0.4076, 0.5435, 0.676, 1.0885, 2.662, 2.662, 2.7381, 2.9781, 3.1589, 4.1746, 4.3598, 4.8724, 9.5612 and 10.2065.

Model-2: Here five observations are generated from Weibull with parameter $\theta = 0.1$ and $\beta = 3$. The remaining ten observations from exponential distribution with parameter $\theta = 0.1$. The ordered observations are 0.7418, 1.3926, 1.4866, 1.5082, 1.5279, 2.1699, 3.0111, 3.1058, 3.4249, 5.6212, 6.5393, 9.1629, 10.2165, 22.0727 and 32.1888.

The identification is done as follows we evaluate for each fixed *r* the maximum likelihood equation $\hat{L}_r$, and then consider $\hat{r}$ being that value of *r* for which likelihood is maximum. The estimates are presented in table(4.4.1) and (4.4.2) for model-1 and model-2 respectively. The *SIC*(0) under Model-1 and Model-2 are 74.22128 and 93.55538 respectively. *BIC* and *HQ* are also found for both the models with the following values.

Table 4.4.1. The Likelihood, parameter estimates and information criterion for Model-1.

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{\beta}$	L	SIC	BIC	HQ
1	11.9537	0.171279	1.29631	-30.204	68.53215	30.47481	33.19269
2	5.92071	0.123884	1.441	-28.4621	65.04835	28.73291	31.45079
3	4.63278	0.081892	1.62678	-26.5204	61.16495	26.79121	29.50909
4	3.87879	0.050043	1.84369	-24.4484	57.02095	24.71921	27.43709
5	**2.31486**	**0.039079**	**1.92499**	**-23.6473**	**55.41875**	**23.91811**	**26.63599**
6	0.864944	0.074318	1.55034	-26.8847	61.89355	27.15551	29.87339
7	0.615074	0.070581	1.53585	-27.5234	63.17095	27.79421	30.51209
8	0.48757	0.061317	1.56115	-27.6898	63.50375	27.96061	30.67849
9	0.400866	0.052183	1.58932	-27.7814	63.68695	28.05221	30.77009
10	0.336613	0.042551	1.62689	-27.7629	63.64995	28.03371	30.75159
11	0.293833	0.042395	1.56627	-28.4133	64.95075	28.68411	31.40199
12	0.260573	0.03771	1.53975	-28.8061	65.73635	29.07691	31.79479
13	0.235881	0.031794	1.50512	-29.215	66.55415	29.48581	32.20369

Table 4.4.2. The Likelihood, parameter estimates and information criterion for Model-2.

r	$\hat{\beta}$	$\hat{\theta}$	L	SIC	BIC	HQ
2	6.25818	0.136196	-41.0662	87.5485	41.15647	42.06243
3	6.25055	0.12443	-38.5097	82.4355	38.59997	39.50593
4	6.32895	0.112701	-35.8889	77.1939	35.97917	36.88513
5	**6.44645**	**0.101003**	**-33.1883**	**71.7927**	**33.27857**	**34.18453**
6	4.47787	0.099347	-33.5527	72.5215	33.64297	34.54893
7	3.25159	0.096469	-34.7522	74.9205	34.84247	35.74843
8	3.03171	0.086918	-34.0851	73.5863	34.17537	35.08133
9	2.8828	0.07774	-33.4219	72.2599	33.51217	34.41813
10	2.28964	0.078485	-35.5625	76.5411	35.65277	36.55873
11	2.06388	0.073388	-36.518	78.4521	36.60827	37.51423
12	1.80399	0.072127	-38.2551	81.9263	38.34537	39.25133
13	1.68415	0.067832	-39.2468	83.9097	39.33707	40.24303

One can observe that the likelihood is maximum and $\min_{1 \le r \le n-1} BIC(r)$ $\min_{1 \le r \le n-1} SIC(r) = SIC(5) < SIC(0)$, and $\min_{1 \le r \le n-1} HQ(r)$ corresponds to $r = 5$, which was expected. The corresponding estimates of the parameters are shown in the tables (4.4.1) and (4.4.2). The graphical representations of the likelihood plot are given in figure (4.4.1) and (4.4.2).

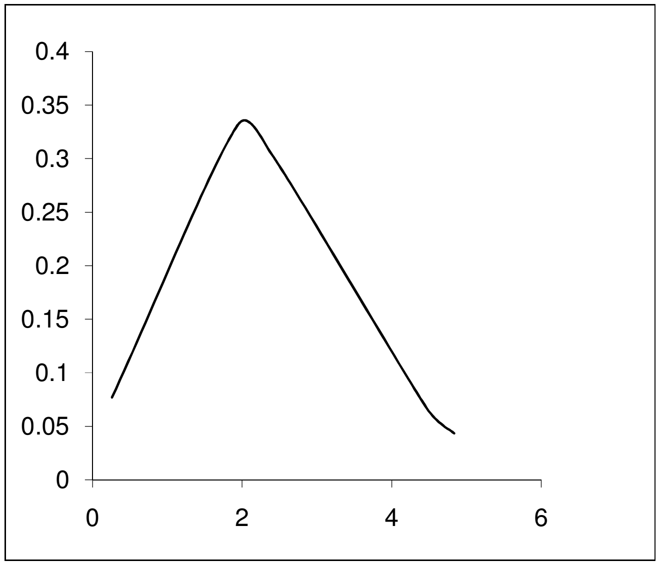


Fig. 4.4.1. The likelihood plot for Model-1

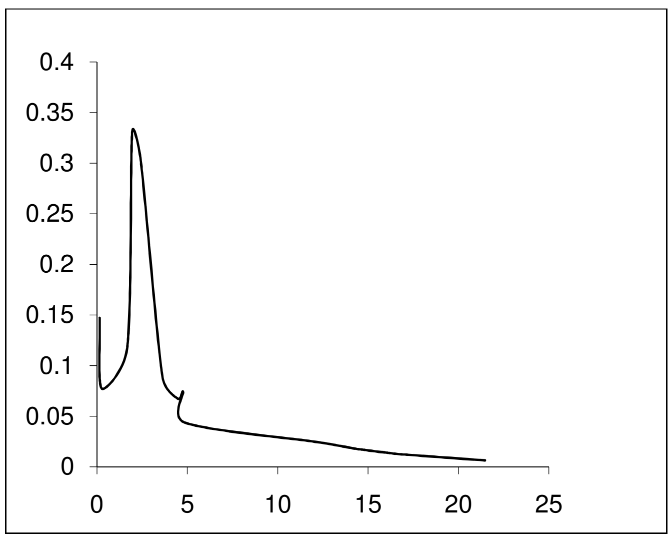


Fig. 4.4.2. The likelihood plot for Model-2

4.5 Data Example:

The example is based on Vanmann's (1991) data on drying of woods under different experiments and schedules. It is the example given in appendix, numbered E-3 S-1.

Under model-1 **:** The computed value $SIC(0)$ = 133.2468 > $SIC(9)$ = min SIC (r) = 98.46836. Also the likelihood is maximum for $\hat{r} = 9$. The corresponding estimates of the parameters are $\hat{\phi}$ =1.40087, $\hat{\beta}$ = 1.96982 and $\hat{\theta}$= 0.015968 as given in the table (4.5.1) below.

Under Model-2: The computed value $SIC(0)$ = 130.3241 and > SIC (13) = min SIC (r) = 125.7627. Hence value of $\hat{r}$ = 13. Similarly other information criteria and likelihood function gives us the same result. The estimates of the parameters are given in table (4.5.2).

Table 4.5.1. Estimates of parameters, likelihood, information criterion under M_1.

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{\beta}$	L	SIC	BIC	HQ
1	17.0411	0.15391	1.1227	-58.6091	126.7524	58.67531	59.76537
2	6.62578	0.125022	1.20295	-56.8636	123.2614	56.92981	58.01987
3	5.61105	0.092835	1.32332	-54.588	118.7102	54.65421	55.74427
4	5.13406	0.06591	1.46092	-52.1563	113.8468	52.22251	53.31257
5	3.87184	0.050274	1.5623	-50.4173	110.3688	50.48351	51.57357
6	3.14867	0.036469	1.68353	-48.5128	106.5598	48.57901	49.66907
7	2.24189	0.029789	1.74933	-47.4397	104.4136	47.50591	48.59597
8	1.77073	0.021963	1.85802	-45.9382	101.4106	46.00441	47.09447
9	**1.40087**	**0.015968**	**1.96982**	**-44.4671**	**98.46836**	**44.53331**	**45.62337**
10	0.789435	0.029368	1.67797	-47.7531	105.0404	47.81931	48.90937
11	0.583143	0.032049	1.61366	-49.0733	107.6808	49.13951	50.22957
12	0.458334	0.036603	1.53082	-50.6362	110.8066	50.70241	51.79247
13	0.383572	0.037393	1.49579	-51.5492	112.6326	51.61541	52.70547
14	0.332632	0.040501	1.43709	-52.7273	114.9888	52.79351	53.88357
15	0.296617	0.041975	1.39664	-53.6335	116.8012	53.69971	54.78977
16	0.271311	0.04458	1.34706	-54.6304	118.795	54.69661	55.78667
17	0.252993	0.047369	1.29805	-55.6037	120.7416	55.66991	56.75997
18	0.240966	0.051993	1.23932	-56.7067	122.9476	56.77291	57.86297
19	0.230694	0.054786	1.19563	-57.6047	124.7436	57.67091	58.76097
20	0.221885	0.055836	1.16131	-58.3575	126.2492	58.42371	59.51377
21	0.215411	0.05678	1.12541	-59.1278	127.7898	59.19401	60.28407
22	0.210003	0.055551	1.09305	-59.8413	129.2168	59.90751	60.99757

Table 4.5.2. Estimates of parameters, likelihood, information criterion under M_2.

r	$\hat{\beta}$	$\hat{\theta}$	L	SIC	BIC	HQ
2	2.45514	4.85387	-65.4516	137.2593	65.51781	67.76414
3	2.14020	4.84449	-64.8678	136.0917	64.93401	67.18034
4	1.94013	4.83674	-64.2642	134.8845	64.33041	66.57674
5	1.84068	4.83136	-63.5878	133.5317	63.65401	65.90034
6	1.76655	4.82876	-62.9371	132.2303	63.00331	65.24964
7	1.73299	4.83517	-62.2321	130.8203	62.29831	64.54464
8	1.70401	4.84410	-61.5473	129.4507	61.61351	63.85984
9	1.68347	4.85958	-60.8549	128.0659	60.92111	63.16744
10	1.60145	4.97633	-60.3106	126.9773	60.37681	62.62314
11	1.53127	5.07469	-59.8869	126.1299	59.95311	62.19944
12	1.43583	5.16651	-59.7701	125.8963	59.83631	62.08264
13	**1.36456**	**5.22507**	**-59.7033**	**125.7627**	**59.76951**	**62.01584**
14	1.28113	5.24744	-59.9096	126.1753	59.97581	62.22214
15	1.21049	5.23142	-60.1883	126.7327	60.25451	62.50084
16	1.13620	5.16086	-60.6704	127.6969	60.73661	62.98294
17	1.06222	5.02958	-61.3154	128.9869	61.38161	63.62794
19	0.90209	4.52081	-63.2469	132.8499	63.31311	65.55944
20	0.82158	4.16407	-64.5197	135.3955	64.58591	66.83224
21	0.72684	3.69465	-66.283	138.9221	66.34921	68.59554
22	0.61228	3.11046	-68.9122	144.1805	68.97841	71.22474

We can observe that, for mixture of two different distributions, we do not get same number of inliers. Now the next problem is to decide which of the model discussed above is better ?

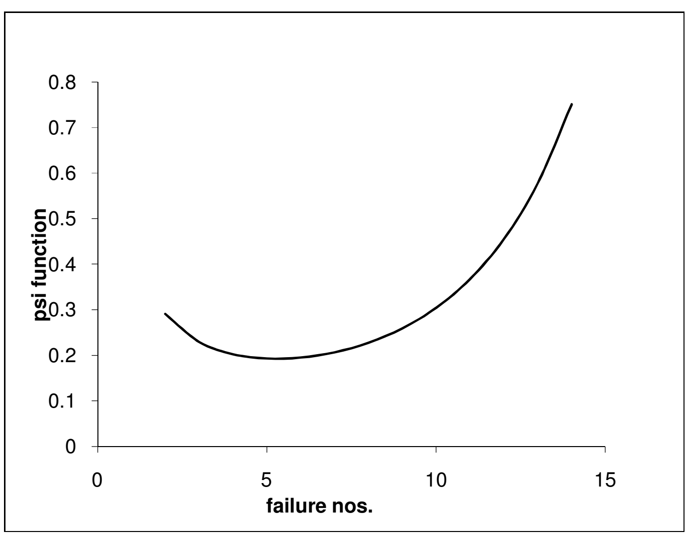


Fig. 4.4.3. The graph of $P[X_{(i)} < X_{(i+1)}]$ under model-1

4.6 Inlier detection using conditional distribution of total lives

This test makes use of basic properties of Poisson process. If one observes Poisson process for a fixed time T and if say n events occur in [$0,T$] at times $0 \le x_{(1)} \le x_{(2)} ..... \le x_{(n)} \le T$ then these times can be considered as ordered observation on a random variable uniformly distributed over [$0,T$]. Let $x_{(i)} =$ life time of i^{th} ordered unit. Then

$$P\Big[x_{(1)} \le X_{(1)} \le x_{(1)} + \Delta x_{(1)}, x_{(2)} \le X_{(2)} \le x_{(2)} + \Delta x_{(2)} .... x_{(n)} \le X_{(n)} \le x_{(n)} + \Delta x_{(n)}$$

$$| n \ events occur(0,T) \ \Big] =$$

$$= \lambda^n e^{-\lambda T} \prod_{i=1}^{n} \frac{\Delta x_{(i)}}{\left(\left[\lambda T\right]^n e^{-\lambda T} / n!\right)}$$

$$= \frac{n!}{T^n} \prod_{i=1}^{n} \Delta x_{(i)}, \qquad x_{(1)} \le x_{(2)} ..... \le x_{(n)} \tag{4.6.1}$$

For large value of n, $\overline{x}$ is approximately normal with mean $\frac{T}{2}$ and variance $\frac{T^2}{12n}$. It can be used to test, for large sample, whether or not the data is drawn from Poisson process. One can also show that if one observes a Poisson Process until exactly n events occur, then (n-1) r.v. can be considered as uniformly distributed over $\left(0, x_{(n)}\right)$.

In context of life testing if the failed items are not placed then all we need to do is to use total lives S_i where $S_i = \sum_{j=1}^{i} D_j$ and $D_i = (n-i+1)\left[x_{(i)} - x_{(i-1)}\right]$. Here $D_i^{'}s$ are known as normalized spacing. If S_i is the total life observed in getting the i^{th} failure then $S_1 \le S_2 ...... \le S_n$.

$$S_r = X_{(1)} + X_{(2)} + ... + (n-r+1) X_{(r)} = \sum_{i=1}^{r} D_i, \quad r = 1,2.....n. \tag{4.6.3}$$

Here also one can show that the total lives $S_1, S_2......, S_n$ can be considered as being drawn from a density function which is uniform over $(0, T)$. If the life test ends as soon as the first *n* failures occur, then the (*n*-1) r.v. $S_1, S_2......, S_{n-1}$ can be considered as being drawn from a density function which is uniform over $(0, S_n)$.

The fact that the conditional distribution of total lives is uniform over suitable interval makes it quiet evident that one has a good tool for detecting whether the failure rate is indeed constant. Thus the contamination of a purely exponential distribution by early failure would manifest itself in the pronounced tendency to get too many clustering together in the early part of total life thus violating uniformity. If the failure rate changes, for example, it increases with time then this should result in a tendency for failures to cluster together as time goes on, again violating uniformity. If the amount of failure data observed is quiet small, then we can expect large changes from exponentiality. Otherwise one can use a chi-square to detect whether the conditional distribution of times to failure or total lives deviate excessively from being normal.

4.6.1 A test for abnormally early failures (inliers)

Suppose that $x_{(1)} \le x_{(2)}...... \le x_{(n)}$ are the *n* ordered failures. If all the $x_{(i)}$ are drawn from a common exponential then S_1 the total life in $\left[0, x_{(1)}\right]$ and $S_n - S_1$, the total life in $\left[x_{(1)}, x_{(n)}\right]$ are distributed independently of each other, where $\frac{2S_1}{\theta} \sim \chi^2_{(2)}$ and $\frac{2(S_n - S_1)}{\theta} \sim \chi^2_{(2[n-1])}$ degrees of freedom each. Hence the ratio

$$R = \frac{(n-1)S_1}{S_n - S_1} \sim F_{(2,2n-2)}. \tag{4.6.4}$$

If the ratio is too small then we assert that $x_{(1)}$ is abnormally small. More precisely if α is the level of significance, we will say $x_{(1)}$ is an inlier if

$$R = \frac{(n-1)S_1}{S_n - S_1} < F_{(2,2n-2),\alpha}. \tag{4.6.5}$$

Suppose one wants to detect $x_{(1)}$ and $x_{(2)}$ are inliers and if all the $x_{(i)}$ are drawn from a common exponential then S_2, the total life in $\left[0, x_{(2)}\right]$ and $S_n - S_2$, the total life in $\left[x_{(1)}, x_{(n)}\right]$ are distributed independently of each other.

$$R = \frac{(2n-2)S_2}{(S_n - S_2)/2} \sim \quad F\,(4, 4n-4) \tag{4.6.6}$$

If this ratio R is too small then we can conclude $x_{(1)}$ and $x_{(2)}$ are inliers. One can continue in similar manner, to detect whether $x_{(1)}, x_{(2)} \ldots\ldots x_{(r)}$ are inliers, where $r = 3, 4, \ldots\ldots n$ till we get first ratio which is greater than tabulated value. Hence at this point one can conclude r observations, till which the hypothesis is accepted, are inliers and rest of the observations are from target population.

4.7 Predictive approach to inlier model detection

The use of predictive distributions has been recognized as the correct Bayesian approach to model determination. In particular, Box(1980) notes the complementary roles of posterior and predictive distributions stating that posterior is used for the "estimation of parameters conditional on the adequacy of the model" whereas the predictive distribution is used for "criticism of the entertained model in the light of the current data". In examining two models, it is clear that the predictive distributions will be comparable whereas the posterior will not.

In this case there are n models, such as model with number of inliers r = 0, 1, 2….n-1. M_1 is considered model with 0 inliers. M_2 can be considered model with r inliers and (n-r) target observations. The procedure is as follows:

Let model M_1 assume that the data X are samples from independent random variables having a target exponential distribution with density

$$f(x|\theta)=\theta e^{-\theta x}, \qquad \theta>0,\ x>0. \tag{4.7.1}$$

The model M_2 assume that there are two distinct labels so that data $X=(X_1,X_2)$ where X_1 and X_2 are sampled from independent random variables having inliers and target exponential distribution, having n_1 = r and n_2 = (n-r) observations respectively, with density function as

$$f(x|\theta_i)=\theta_i e^{-\theta_i x}, \qquad \theta_i>0,\ x>0,\ i=1,2. \tag{4.7.2}$$

where $\theta_1=\phi$ the parameter of inliers distribution and $\theta_2=\theta$ the parameter of target distribution. If assumption regarding the vague prior density of the form under M_1 is $g(\theta)\propto\frac{1}{\theta}$. The likelihood under model M_1 is as follows

$$\begin{aligned} L(\underline{X},M_1)&=\prod_{i=1}^{n} f(x_i|\theta,M_1) \\ &=\theta^n e^{-\theta\sum_{i=1}^{n} x_i} \end{aligned}$$

Then predictive density of observation x under M_1 is given by

$$f(x|X,M_1)=\frac{\int L(x,\theta,M_1)L(\underline{X},\theta,M_1)g(\theta)d\theta}{\int L(\underline{X},\theta,M_1)g(\theta)d\theta} \tag{4.7.3}$$

The model M_2 assume that there are two distinct labels so that data X=(X_1, X_2) where X_i are sampled from independent random variables having a distinct exponential distribution with density

$$f(x|\theta_i)=\theta_i e^{-\theta_i x}, \qquad i=1,2,\ \ \theta_i>0,\ x>0. \tag{4.7.4}$$

The vague prior densities of the form g(θ_i) $\propto$ θ_i^{-1} for both the parameters are assumed, then the respective predictive densities under M_1 and M_2 are

$$f\left(x \mid X, M_1\right) = n(n\bar{x})^n / \left(n\bar{x} + x\right)^{n+1} \tag{4.7.5}$$

$$f\left(x \mid X, M_2\right) = n_i(n_i\bar{x}_i)^{n_i} / \left(n_i\bar{x}_i + x\right)^{n_i+1}, \quad i = 1,2 \tag{4.7.6}$$

where

$$\bar{x} = n^{-1}\left(n_1\bar{x}_1 + n_2\bar{x}_2\right) \quad \text{and} \quad \bar{x}_i = \frac{\sum_{i=1}^{n_i} x_i}{n_i}, \quad i = 1,2. \tag{4.7.7}$$

The prior density yields the optimal estimate of the density, in the frequency sense, among all estimates that are invariant with regards to transformation of scale using Kullback- Leibler measure of divergence.

The Predictive sample reuse (PSR) quasi–bayes criterion chooses the Larger of

$$L_1 = \frac{\prod_{i=1}^{2}\prod_{j=1}^{n_i}(n-1)\left(n\bar{x} - x_{ij}\right)^{n-1}}{n\bar{x}} \qquad for\ x_{ij} > 0 \tag{4.7.8}$$

and

$$L_2 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n_i - 1)\left(n_i\bar{x}_i - x_{ij}\right)^{n_i-1}}{\left(n_i\bar{x}_i\right)^{n_i}} \quad for\ x_{ij} > 0 \tag{4.7.9}$$

The Predictive sample reuse (PSR) quasi–bayes criterion used by Geisser and Eddy (1979) chooses the model with Larger of

$$\hat{L}_1 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n-1)}{\left(n\bar{x} - x_{ij}\right)}\exp\left(-\frac{(n-1)x_{ij}}{\left(n\bar{x} - x_{ij}\right)}\right) \qquad for\ x_{ij} > 0 \tag{4.7.10}$$

and

$$\hat{L}_2 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n_i - 1)}{\left(n_i\bar{x}_i - x_{ij}\right)^{n_i-1}}\exp\left(-\frac{(n_i - 1)x_{ij}}{\left(n_i\bar{x}_i - x_{ij}\right)^{n_i-1}}\right) \quad for\ x_{ij} > 0 \tag{4.7.11}$$

Above mentioned both the criterions are asymptotically equivalent to Akaike's criterion. One can use any of the above given criteria to obtain number of inliers in a given set of data.

4.8 Numerical illustration

The data represents ozone concentration in ppb monitored from morning 8 a.m. to evening 8 p.m. at express highway of Anand in the month of July on hourly basis. The data is collected by Dr. Sukalyan Chakraborty as a part of air pollution status monitoring of Anand district for his research. The observations arranged in increasing order of their magnitude are 14.00, 14.50, 15.00, 15.00, 17.00, 17.00, 19.00, 21.00, 21.80, 22.30, 23.00, 23.20 and 24.00. In table (4.8.1), *r* represents number of inliers observations to be considered. Level of significance is taken as 2.5 %.

Table 4.8.1. Inlier detection using Likelihood and Conditional method

r	Likelihood	$\hat{L}_2$	For Conditional Method				Conclusion
			D_i	S_i	Ratio	F-tab	
1	-51.2223	---------	182	182.0	1.5888	4.318725	Accept
2	-51.1812	5.7383E-23	174	356.0	1.6308	3.066233	Accept
3	-51.1434	5.9886E-23	165	521.0	1.6769	2.589498	Accept
4	-51.0982	6.3160E-23	150	671.0	1.7047	2.327027	Accept
5	-51.0853	6.3908E-23	153	824.0	1.7996	2.157011	Accept
6	**-51.0636**	**6.5672E-23**	**136**	**960.0**	**1.8773**	**2.036182**	**Accept**
7	-51.0712	6.5008E-23	133	1093.0	2.0208	1.944986	Reject
8	-51.1041	6.2351E-23	126	1219.0	2.2567	1.873191	
9	-51.1387	5.9862E-23	109	1328.0	2.5818	1.814874	
10	-51.1712	5.7747E-23	89.2	1417.2	3.0499	1.766351	
11	-51.2051	5.5642E-23	69.0	1486.2	3.8383	1.725199	
12	-51.2345	3.5233E-21	46.4	1532.6	5.3215	1.689750	
13	-52.2672	----------	24.0	1556.6			

For conditional method null hypothesis is rejected when r = 7 implies that number of inlier in the data set is 6 as shown in table (4.8.1). The likelihood is also maximum at *r = 6*. Using Predictive Method we have obtained $\hat{L}_1 = 5.21825\text{E}-23$ and maximum $\hat{L}_2 = 6.5672\text{E}-23$ corresponds to *r = 6*.

4.9 Goodness of fit

The problem of testing of goodness of fit to test whether the sample data is taken from modified mixture Weibull distribution against they are taken from single exponential or Weibull distribution is discussed in this section.

4.9.1 To test whether target observations are from Exponential

Our first test is

H_0 : the sample is from single population with exponential distribution i.e. $f(x,\theta)$

H_1 : the sample is from population with Modified Weibull distribution,

In terms of the MLE, the likelihood ratio test statistics for testing H_0 against H_1 is

$$\Lambda = \frac{L(\theta \mid H_0)}{L(\phi,\theta,\beta \mid H_1)} \tag{4.9.1}$$

$$\ln\Lambda = n\ln\theta - \theta\sum_{i=1}^{n} x_{(i)} - r\left[\ln\phi + \ln\beta_0\right] - (n-r)\left[\ln\theta + \ln\beta_1\right] - \\ (\beta_0 - 1)\sum_{i=1}^{r}\ln x_{(i)} - (\beta_1 - 1)\sum_{i=r}^{n}\ln x_{(i)} + \phi\sum_{i=1}^{r} x_{(i)}^{\beta_0} + \theta\sum_{i=r+1}^{n} x_{(i)}^{\beta_1} \tag{4.9.2}$$

Under null hypothesis $Y_L = -2\ln(\Lambda) \sim \chi_4^2$. Reject H_0 for appropriate value of level of significance when $Y_L > \chi_{4,\alpha}^2$.

4.9.2 To test whether all observations are from single Weibull against they are from mixture of two (inliers and target) Weibull distributions.

Our second test is

$H_0 : p = 1$ the sample is from single (target) population from Weibull distribution with parameters $\beta \neq 1$ and $\theta > 0$.

$H_1 : p < 1$ the population distribution is Modified Weibull with parameters $\beta \neq 1$, $\phi > 0$ and $\theta > 0$.

In terms of the MLE, the likelihood ratio test statistics for testing H_0 against H_1, as used in test 1, is

$$\ln \Lambda = n\left[\ln \theta + \ln \beta\right] + (\beta - 1)\sum_{i=1}^{n} \ln x_{(i)} - \theta \sum_{i=1}^{n} x_{(i)}^{\beta} - r\left[\ln \phi + \ln \beta_0\right] - (n-r)\left[\ln \theta + \ln \beta_1\right]$$
$$- (\beta_0 - 1)\sum_{i=1}^{r} \ln x_{(i)} - (\beta_1 - 1)\sum_{i=r}^{n} \ln x_{(i)} + \phi \sum_{i=1}^{r} x_{(i)}^{\beta_0} + \theta \sum_{i=r+1}^{n} x_{(i)}^{\beta_1}$$

(4.9.3)

Under null hypothesis $Y_L = -2\ln(\Lambda) \sim \chi_2^2$, then reject H_0 for appropriate value of level of significance when $Y_L > \chi_{2,\alpha}^2$.

4.9.3 Sequential Probability ratio test (SPRT)

SPRT is used to find number of inliers in given data set for both the models as shown in the following sub sections.

Case 1 : SPRT for model-1

To test whether inliers and target population is from single Weibull distribtion against they are from two different Weibull population, i.e with reference to section (1.5). The SPRT test is given as follows

H_0: Sample observations are taken from inlier population with interest parameter $\xi = \phi$.

H_1: Sample observations are taken from target population with interest parameter $\xi = \theta$..

and likelihood ratio λ_m is given by $\lambda_m = \frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln \lambda_m = \sum_{i=1}^{m} \ln \frac{f\left(x_{(i)}, \theta\right)}{g\left(x_{(i)}, \phi\right)} =$$

$$= m\left(\ln \theta + \ln \beta_1 - \ln \phi - \ln \beta_0\right) + \left[\beta_1 - \beta_0\right] \sum_{i=1}^{m} \ln x_{(i)} - \theta \sum_{i=1}^{m} x_{(i)}^{\beta_1} + \phi \sum_{i=1}^{m} x_{(i)}^{\beta_0} \qquad m = 1, 2, \ldots. n$$

(4.9.5)

For deciding number of inliers *r*, first arrange the observations in ascending order and then continue to take likelihood ratio for *m* = 1, 2…. by including observations one by one till we reject H_0. That is

if $\sum_{i=1}^{m} z_{(i)} \leq \ln B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{m} z_{(i)} \geq \ln A$ reject H_0 and stop.

The corresponding m represents the first observation from $f\left(x_{(i)}, \theta\right)$ and number of inliers $\hat{r} = m - 1$. Also

$$B = \frac{\gamma}{1-\alpha} \qquad A = \frac{1-\gamma}{\alpha} \tag{4.9.6}$$

where α represents probability of type I error and γ represents probability of type II error. Arrange $X_{(1)} \leq X_{(2)} \leq \ldots\ldots X_{(n)}$ and apply SPRT process till the hypothesis H_0 is rejected.

Case 2: SPRT for model-2

To test whether observations follow Weibull distribution against they follow exponential distribution. The SPRT test is given as follows:

H_0 : Inlier observations are taken from Weibull population

H_1 : Inlier observations from Weibull and target from exponential population

and likelihood ratio λ_m is given by $\lambda_m = \dfrac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln \lambda_m = \sum_{i=1}^{m} \ln \frac{f\left(x_{(i)}, \theta\right)}{g\left(x_{(i)}, \phi\right)} = -m \ln \beta_0 - \left(\beta_0 - 1\right) \sum_{i=1}^{m} \ln x_{(i)} + \theta \left(\sum_{i=1}^{m} x_{(i)}^{\beta} - \sum_{i=1}^{m} x_{(i)} \right)$$

$$\text{where} \quad m = 1, 2, \ldots. n \tag{4.9.7}$$

For deciding number of inliers *r*, first arrange the observations in ascending order and then we continue to take likelihood ratio for *m*= 1, 2….*n* by including observations one by one till we reject H_0. Arrange $X_{(1)} \le X_{(2)} \le \ldots\ldots X_{(n)}$ and apply SPRT process till the hypothesis H_0 is rejected.

Test criteria for rejection of H_0, using $\ln \lambda_m$ as defined for case 1 and case 2 in equations (4.9.6) and (4.9.7) is to reject H_0, if

$$\ln \lambda_m > \ln A \tag{4.9.8}$$

Corresponding value of *m* for which H_0 was accepted last becomes number of inliers *r.*

4.10 Conclusion

The Akaike information criterion is a measure of the relative goodness of fit of a statistical model.. It can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking between accuracy and complexity of the model.

Given a data set, several candidate models may be ranked according to their *AIC* values. From the *AIC* values one may also infer that e.g. the top two models are roughly in a tie and the rest are far worse. Thus, *AIC* provides a means for comparison among models—a tool for model selection. In general $AIC = 2k - 2 \ln L$, where *k* is the number of parameters in the statistical model, and *L* is the maximized value of the likelihood function for the estimated model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters.

To compare above two models, defined in section (4.3.1) and (4.3.2), obtained value of AIC for Model-1 is 94.9342 and for Model-2 is 123.4066. Clearly we can observe Model-1 is better than Model-2. i.e. Model representing inliers and target observations as Weibull distribution with different scale parameters is better. For same example discussed in section (4.5), the Pareto distribution had also been applied in chapter 2. Hence comparing Weibull against Pareto model, it was noted that *AIC* for Weibull distribution is 127.7126 > AIC for Pareto distribution is 59.17455. Hence one can conclude for that example Pareto model is better than Weibull model. The Pareto distribution is a power-tailed distribution which is a special case of a heavy-tailed distribution whose tails go to zero more slowly than exponential. In particular, in the cases where initial defects are present causing early failures, the Pareto distribution is found adequate to model such phenomenon.

Above result is supported by Jian-ming Mo and Zong-Fang (2008) who compared the sensitivity of aggregate operational value-at-risk in the Pareto distribution with that in the Weibull distribution to select an optimal model from the loss severity distributions of approximate goodness-of-fit. After the aggregate operational value-at-risk is obtained, the sensitivities of aggregate operational value-at-risk are compared when the loss severity distribution are respectively the Pareto and Weibull. The authors have shown that the sensitivity of aggregate operational value-at-risk with the Pareto distribution is far better than that with the Weibull distribution.

Another paper that discussed the comparision of Pareto and Weibull model was by Li-Hua Lai, Khoo, Murlidharan and Xie (2007) and Pei-Hsuan Wu (2008) and Wo-Chiang Lee (2009) have shown that using extreme value theory, generalized

Pareto distribution (GPD) fits the heavy-tailed distribution better than the lognormal, gamma, Weibull and normal distributions. In an empirical study, they determine the thresholds of GPD through mean excess plot and Hill plot.

Chapter 5

Inliers estimation in complete mixtures

5.1 Introduction

Finite mixtures distributions have provided a mathematical-based approach to the statistical modeling of a wide variety of random phenomena. Because of their usefulness as an extremely flexible method of modeling, finite mixture models have continued to receive increasing attention over the years, from both practical and theoretical points of view. Indeed, in the past decade the extent and the potential of applications of finite mixture models have widened considerably. Fields in which mixture models have been successfully applied include astronomy, biology, genetics, medicine, psychiatry, economics, engineering, and marketing, among many other fields in the biological, physical and social sciences.

Mixture distributions have been extensively used in a wide variety of important practical situations where data can be viewed as arising from two or more populations mixed in varying proportions. Mixture of distributions refers to the situation in which i^{th} distribution out of k underlying distribution is chosen with probability p_i, i=1,2,….k. Mixture distribution having k=2 components are extensively studied in literature. For example a probability model for the life of an electronic product can be described as the mixture of two uni-model distribution, one representing the life of inliers and other for target observations. A mixture model is able to model quite complex distributions through an appropriate choice of its components to represent accurately the local areas of support of the true distribution. The problem of central interest arises when data are not available for each distribution separately, but only for the overall mixture distribution. Often such situations arise because it is impossible to observe some underlying variable which splits the observations into groups then only the combined distribution can be studied. In these circumstances, interest often focuses on estimating the mixing proportions and on estimation of the parameters in the conditional distributions. There is a remarkable variety of estimation methods that have been applied to finite mixture problems such as graphical methods, the method of moments, maximum likelihood, minimum chi-square, least squares approaches and Bayesian approaches.

Decomposing a finite mixture of a distribution is a very difficult problem as it can be observed looking at the solution based on method of moments put forward by Karl Pearson (1894) in the case of a mixture of two univariate normal distributions with unequal variances. However, Tan and Chang (1972) have shown that the method of moments is inferior to likelihood estimation for this problem.

Finite mixture models have been broadly developed and widely applied to classification, clustering, density estimation and pattern recognition problems, as shown by Titterington, Smith and Markov (1985), McLachlan and Basord (1988), Lindsay (1995), B¨ohning (1999) and Peel (2000), and the references therein. With the growing advances of computational methods, especially for the development of Markov chain Monte Carlo (MCMC) techniques, many works are also devoted to Bayesian mixture modeling issues, including and Diebolt and Robert (1994), Escobar and West(1995), Richardson and Green (1997) and Stephens (2000), among others.

Because of their usefulness as an extremely flexible method of modeling, finite mixture models have continued to receive increasing attention over the years, both from a practical and theoretical point of view. For multivariate data of a continuous nature, attention has been focused on the use of multivariate normal components because of their computational convenience. They can be easily fitted iteratively by maximum likelihood (ML) via the expectation maximization (EM) algorithm of Dempster, Lai, Khoo, Murlidharan and Xierd and Rubin (1977) and McLachlan and Krishnan (1997). By adopting some parametric form for the density function in each underlying group, likelihood can be formed in terms of mixture distribution and unknown parameter estimated by consideration of the likelihood. The likelihood approach to fitting of mixture models in particular normal mixtures has been utilized by several authors, Dick and Bowden (1973) and O'Neill (1978) .

In the last two decades, the skew normal distribution has been shown beneficial in dealing with asymmetric data in various theoretic and applied problems. Authors took up the problem of analyzing a mixture of skew normal distributions from the likelihood-based and Bayesian perspectives, respectively. Computational techniques using EM-type algorithms are employed for iteratively computing maximum likelihood estimates done by Lin, Lee and Yens (2007).

Andersen(1996) introduced a modification of the mixture of distributions model based on microstructure arguments. Based on a small sample of five stocks, he infers that this modified mixture of distributions (MMD) model adequately captures the joint behavior of trading volume and volatility. He re-examine this cLai, Khoo, Murlidharan and Xiem using a larger sample of twenty-two stocks and two sample periods.

Chen and Kalbfleisch (2005) and Chen et al. (2001, 2002) suggest a modification of the likelihood by incorporating a penalty term that forces certain estimates away from the boundary of the parameter space. The likelihood ratio statistic based on the modified estimators is shown, in many instances, to yield relatively simpler limiting distributions and hence simpler tests.

Finite mixture models belong to a class of non-regular models and, as a consequence, many classical asymptotic results do not apply. Many researchers have tried to understand the large sample properties related to the analysis of finite mixture models. Hartigan (1985) first demonstrated the peculiar behavior of the likelihood ratio statistic for mixture models. Ghosh and Sen (1985) obtained the limiting distribution under a separation condition. The separation condition turned out to be unnecessary, which was shown by Chernoff and Lander (1995) for binomial mixtures, and in general by Chen and Chen (2001, 2002), Dacunha and Gassiat (1999) and others. Even though the large sample behavior of the likelihood ratio statistic under a mixture model is now better understood, its implementation still poses a challenge. The main difficulty involves determining the critical value based on a limiting distribution that involves the supremum of a Gaussian process. Techniques given in Adler (1990) and Sun (1993) may be useful in this respect. An alternative, discussed in McLachlan (1987), Chen (1998), Chen and Chen (2001) and elsewhere, is to use re-sampling methods. Bayesian methods can also be applied in this context as done by Richardson and Green (1997). Additional recent work can be found in McLachlan and Peel (2000), Lo et al. (2001), Garel (2001) and Garel and Goussanou (2002).

A popular way to account for unobserved heterogeneity is to assume that the data are drawn from a finite mixture distribution. A barrier to using finite mixture models is that parameters that could previously be estimated in stages must be estimated jointly because using mixture distributions destroys any additive separability of the log-likelihood function. Arcidiacono and Jones(2002) show, however, that an extension of the EM algorithm reintroduces additive separability, thus allowing one to estimate parameters sequentially during each maximization step. In establishing this result, the author developed a broad class of estimators for mixture models. Returning to the likelihood problem, relative to full information maximum likelihood, the sequential estimator can generate large computational savings with little loss of efficiency.

Mixture models, in which a probability distribution is represented as a linear superposition of component distributions, are widely used in statistical modeling and pattern recognition. One of the key tasks in the application of mixture models is the determination of a suitable number of components. Conventional approaches based on cross-validation are computationally expensive, are wasteful of data, and give noisy estimates for the optimal number of components. A fully Bayesian treatment, based on Markov chain Monte Carlo methods for instance, will return a posterior distribution over the number of components. However, in practical applications it is generally convenient, or even computationally essential, to select a single, most appropriate model. Recently it has been shown, in the context of linear latent variable models, that the use of hierarchical priors governed by continuous hyper-parameters whose values are set by type-II maximum likelihood, can be used to optimize model complexity. Author extends a framework to mixture distributions by considering the classical task of density estimation using mixtures of Gaussians. They show that, by setting the mixing coefficients to maximize the marginal log-likelihood, unwanted components can be suppressed, and the appropriate number of components for the mixture can be determined in a single training run without recourse to cross validation. Their approach uses a variational treatment based on a factorized approximation to the posterior distribution by Corduneanu and Bishop (2001).

Bayesian predictive density functions, which are necessary to obtain bounds for predictive intervals of future order statistics, are obtained when the population density is a finite mixture of general components. Such components include, among others, the Weibull (exponential and Rayleigh as special cases), compound Weibull (three-parameter Burr type XII), Pareto, beta, Gompertz and compound Gompertz distributions. The prior belief of the experimenter is measured by a general distribution that was suggested by AL-Hussaini (2003). Applications to finite mixtures of Weibull and Burr type XII components are illustrated and comparison is made, in the special cases of the exponential and Pareto type II components, with previous results.

Everitt and Bullmore (1999) report on a novel method of identifying brain regions activated by periodic experimental design in functional magnetic resonance imaging data. This involves fitting a mixture distribution with two components to a test statistic estimated at each voxel in an image. The two parameters of this distribution, the proportion of nonactivated voxels (inliers) and the effect size can be estimated using maximum likelihood methods. Standard errors of the parameters can also be estimated. The fitted distribution can be used to derive brain activation maps and two examples are described, one involving a visual stimulation task, the other an auditory stimulation task. The method appears to have some advantages over direct use of the *P*-values corresponding to each voxel's value of the test statistic.

The merits and limitations of parametric and nonparametric methods and the value of historical floods and palaeoflood information are reviewed and discussed. A mixture density estimation procedure based on the Gumbel (EV1) distribution kernel is introduced and a modified maximum likelihood criteria is developed for estimation of model parameters by Guo Shen Lian (2009). Using the recorded data and pre-gauging floods in China and a limited number of simulation experiments, the flood quantiles estimated by the proposed model are compared with those estimated by parametric and nonparametric methods. It is found that the mixture density estimation method can fit real data points more closely than its parametric counterparts, and that it is competitive with the other considered candidates.

5.2 Inliers as mixture model

An inlier in a set of data is an observation or subset of observations not necessarily all zeros, which appears to be inconsistent with the remaining data set. Consider an example where the weights of new born 17 babies (in pounds) in a hospital is noted as 0, 0, 1.2, 1.4, 2, 3.5, 3.8, 4.2, 4.6, 5.5, 5.5, 5.8, 6, 6.2, 6.6, 6.6 and 7. Observation 0 can be considered as child born dead. As we have already seen in chapter 1, by specifying $\delta = 4$, first 6 observations can treated as inliers. The observations which are identified as instantaneous and early failures together are called inliers, introduced first time by Muralidharan and Kale (2002).

Apart from the examples discussed in introduction chapter (chapter 1) the following examples also gives us the idea of inlier generation as a complete mixtures.

1. To study the growth in dog's population, one may observe age of dogs. We can observe there are some puppies with no life, some with short life span while rest of them live the average target life. The observations of no life or short life span can be considered as inliers.

2. In the production of electronic components of air conditioner, some components may fail on installation and therefore have zero life lengths. A component that does not fail on installation will have a life length that is a positive random variable whose distribution may take different forms. We can take component which fail instantaneously or early as inliers. Thus, the overall distribution of lifetimes is a nonstandard mixture.

3. Consider profit earned on a share during a long term. There will be times when we get no profit and times when profit is continuous distribution of positive value. The observation with zero profit and small values of profit can be considered to be as inliers.

4. In a clinical trial laboratory a particular drug is designed and given to certain species of 100 hens so that the new chicks have weight greater than usual. The possible observations may be combination of inliers (i.e. no gain or negligible gain of weight than usual) and target gain in weight.

Inliers can be classified into discordant observations (those which appear "surprising or discrepant" very small to investigator) and contaminants mixture model of the form

$$h(x) = (1-p)g(x) + pf(x)$$

Here one can consider $f(x)$ as our target density function (pdf of interest) and $g(x)$ as inlier density function. The objective is to estimate the proportion $(1-p)$ of observation coming from $g(x)$ which is very small as compared to the observations of $f(x)$. This can be achieved by carrying out a test procedure for $H_0: p = 1$ against $H_1: p < 1$ and decide whether samples are from $g(x)$ or $f(x)$.

5.3 Methods to detect inliers

Over the past years, a variety of methods have been developed for estimating the parameters in finite mixture models. Four of them are widely used in practice and cited in the literature, they are graphical method, method of moments, minimum-distance method, maximum likelihood method and Bayesian method. The method of moments is the earliest method for estimating the parameters in finite mixture models. The estimation procedures for inlier observations are present in the model given below.

5.3.1 Graphical methods based on probability model

It is the easiest way to find whether data is from mixed population. The two most common graphs which can give us idea whether the sample observations are from single population or are they taken from population which is mixture of two populations (one of them represents inliers and other is continuous life time distribution). One can easily identify presence of mixture of two distributions, just looking at the graph.

a. Density function graph

The graph represents mixture of two normal distributions. The graph (5.3.1) represents inlier and target observations taken from $N(7,9)$ and $N(27,9)$, respectively. From the graph we can identify two symmetrical curves such that first curve has mean remarkably less than second curve which can be considered as inlier distribution. Similarly we can have graph representing mixture of more than two distributions. From the graph (5.3.1) one can get rough idea about number of components with approximate mean. The density graph of mixture of inliers and target population for distribution other than normal is discussed in section (5.5).

b. Cumulative distribution function

It is also known normal quantile quantile (Q-Q) plot. This plot can be described as plot of an estimate of $F^{-1}(p)$ against $\varphi^{-1}(p)$, where $F(\cdot)$ is the cumulative distribution function of the mixture distribution and $\varphi(\cdot)$ is that of standard normal. A sample from single normal distribution should produce a linear plot. Refer graph (5.3.2) which indicates the presence of mixture of two distributions (data used is same as above section).

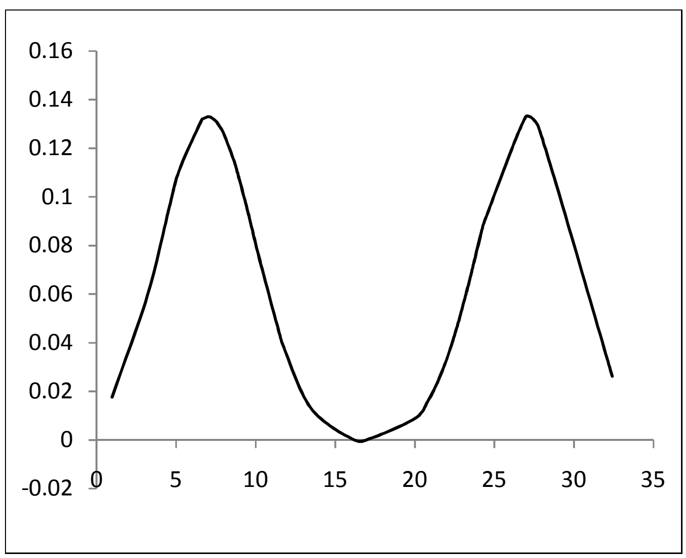


Fig. 5.3.1. Density function of mixture distribution

The graph (5.3.2) indicates deviation from linearity which is the characteristic of certain type of mixture of two populations.

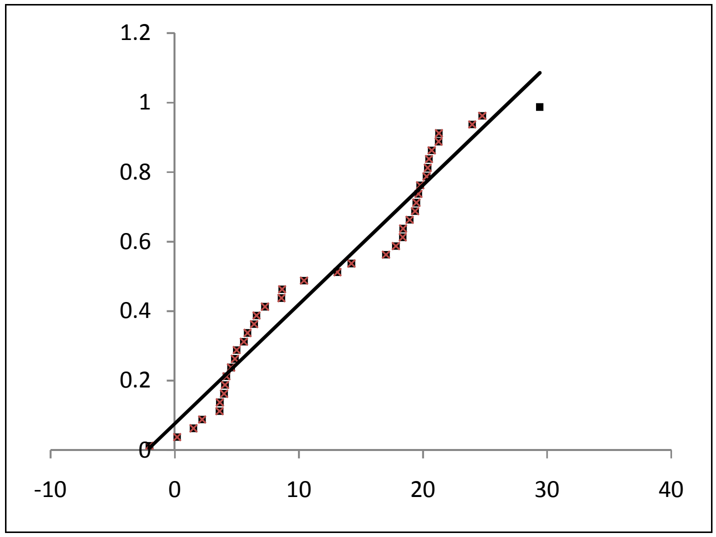
c. 

Fig. 5.3.2. Quantile-Quantile plot

5.3.2 Method of moments

Suppose we have data set with *n* independent observations from a population whose probability model depends on *v* unknown parameter, ξ. Let $\mu(\xi)$ denote vector of *v* functionally independent moments and that *m* denotes the corresponding set of sample moments. The method of moments estimator is the ξ. which satisfies

$$\mu\left(\hat{\xi}\right)=m \qquad (5.3.1)$$

If ξ denotes the mean of mixture distribution of inliers and target distribution, then we will get sample mean which will be a value approximately the average of the above two groups.

There are many problems in using moment estimators, such as

a. Explicit solution of (5.3.1) may not be easy or even possible.
b. The solution to (5.3.1) may not be unique.
c. They may not be asymptotically efficient.

To answer these questions, we proceed with other estimation procedure.

5.3.3 Method of maximum likelihood

The data in the random sample are of the form $X_1 = x_1, X_2 = x_2, \ldots X_n = x_n$, where the distribution of each X is described by a parametric finite mixture density. Most statistical methods will then take their starting point the likelihood function as

$$L_0(\xi) = \prod_{i=1}^{n} P(x_i \mid \xi) = \prod_{i=1}^{n} \left[(1-p) g(x_i) + p f(x_i) \right] \tag{5.3.2}$$

Maximization of $L_0(\xi)$ with respect to ξ, for given data X, yields the maximum likelihood estimates of parameter ξ. Normally the quantity maximized is log-likelihood $\mathcal{L}_0(\xi) = \ln L_0(\xi)$.

Even in mixture models, maximum likelihood approach is very popular because

a. It fits into the philosophy of likelihood-based inference.
b. The existence of attractive asymptotic theory.
c. The estimates are often easy to compute.
d. They are also useful for calculating Bayesian posterior modes.

For inliers mixture model many times the asymptotic theory and computational aspects are not so straight forward. In such case one has to use iterative methods to obtain the estimates of the parameters of inliers mixture distribution.

5.3.4 Minimum-distance method

Another general method for estimating the mixing distribution in finite mixture model is to minimize the distance between the empirical distribution and the mixture distribution or the distance between the kernel density estimation and the mixture density. Titterington et al. (1985) gave a detailed review of the minimum-distance estimators. Maximum likelihood estimator can also be viewed as a special case of minimum-distance estimators, simply because it minimizes the Kullback-Leibler (1951) distance between the empirical distribution and the mixture distribution. Due to the rapid improvement in computing power, finding numerical solutions of a likelihood equation becomes feasible. Likelihood-based inference has enjoyed fast development and plays an important role in the scope of finite mixture models.

5.3.5 Bayesian method

One of the methods for estimating parameter of mixture distribution is the Bayesian method. Let $\ln(x_1, x_2, ...x_n | \theta)$ be the likelihood function of θ. In the framework of the Bayesian approach, one needs to assume that a prior distribution $P(\theta)$ when θ is available. Using Bayes' theorem, we can obtain the posterior density $P(\theta | x_1, x_2, ...x_n)$ which is given by

$$P(\theta | x_1, x_2, ...x_n) \;\; \alpha \;\; \ln(x_1, x_2, ...x_n | \theta) P(\theta) \tag{5.3.3}$$

There are two main reasons why people may be interested in using the Bayesian method in finite mixture models. Firstly, including a suitable prior distribution for θ in the framework of the Bayesian approach may avoid spurious modes when maximizing the log-likelihood function. Secondly, when the posterior distribution for the unknown parameters is available, the Bayesian method can yield valid inference without relying on the asymptotic normality. As warned by McLachlan and Peel (2000), the asymptotic theory of the MLE can apply only when the sample size *n* is very large. Hence the second advantage of the Bayesian method become obvious when the sample size *n* is small.

The iterative methods used for estimation of parameters in above method are discussed in next sub section. There are three well known iterative procedure to estimate the values of parameter of mixture distribution

5.3.6 Expectation Maximization (EM)

EM algorithm is an iterative method to obtain estimates of parameters which are not in an explicit form. EM algorithm works as follows:

Suppose we have to find $\xi = \hat{\xi}$ to maximize the likelihood

$$L(\xi) = f(x \mid \xi), \tag{5.3.4}$$

where *x* is set of "inlier" data. Let y denote a typical "complete" version of *x* and let y(x) denote the set of all possible such *y*. In inlier mixture context of equation (5.3.4) the likelihood of *y* be denoted by $g(y \mid \xi)$. The EM algorithm generates, from some initial approximation, $\xi^{(0)}$, a sequence $\{\xi^{(m)}\}$ of estimates. Each iteration consists of the following double step:

E step : Evaluate $E\left[\ln\{g(y \mid \xi)\} \mid x, \xi^{(m)}\right] = Q\left(\xi, \xi^{(m)}\right)$.

M step: Find $\xi = \xi^{(m+1)}$ to maximize $Q\left(\xi, \xi^{(m)}\right)$.

The Expectation-Maximization algorithm for the finite mixture problem proposed by Dempster, Lai, Khoo, Murlidharan and Xie (2007), and Rubin (1977), popularly known as the EM algorithm, is a broadly applicable approach to the iterative computation of MLE's, useful in a variety of incomplete-data problems, where algorithms such as the Newton-type methods may turn out to be more complicated.

5.3.6.1 Inlier detection in normal distribution using EM algorithm

Let $X = (x_1, x_2, ...x_n)$ be a sample of *n* independent observations from mixture of two inlier and target normal distributions and let $Z = (z_1, z_2, ...z_n)$ be latent variables that determines component from which observation originates. $X_i \mid (z_i = 1)$ follows $N(\mu_1, \sigma_1)$ and $X_i \mid (z_i = 2)$ follows $N(\mu_2, \sigma_2)$, where $P(z_i = 1) = p_1$ and The aim is to estimate the unknown parameters representing the "mixing" values of $\theta = (p, \mu_1, \mu_2, \sigma_1, \sigma_2)$.

The likelihood function is given by

$$L(\theta,x,z)=P(x,z\,|\,\theta)=\prod_{i=1}^{n} I(Z_i=j)\left[p_1 g(x_i)+p_2 f(x_i)\right]$$

where I is an indicator function.

<u>E- step</u>

$$T_{ji}^{(t)}=P\left(z_i=j\,|\,X_i=x_i:\theta^{(t)}\right)$$

$$=\frac{p_j^{(t)} f\left(x;\mu_j^{(t)},\sigma_j^{(t)}\right)}{p_1^{(t)} f\left(x;\mu_1^{(t)},\sigma_1^{(t)}\right)+p_2^{(t)} f\left(x;\mu_2^{(t)},\sigma_2^{(t)}\right)}$$

Thus E-step result in the function

$$Q\left(\theta\,|\,\theta^{(t)}\right)=\sum_{i=1}^{n}\sum_{j=1}^{2} T_{ji}^{(t)}\left[\ln p_j-\frac{1}{2}\ln\sigma_j-\frac{1}{2}\frac{\left(x_i-\mu_j\right)^2}{\sigma_j^2}-\frac{1}{2}\ln(2\pi)\right]$$

and

<u>M-step</u>

The quadratic form $Q\left(\theta\,|\,\theta^{(t)}\right)$ means that determining the maximizing values of θ is relatively straight forward. Firstly note that $p,\left(\mu_1,\sigma_1\right)\&\left(\mu_2,\sigma_2\right)$ may be all maximized independently of each other since they all appear in separate linear terms. The estimates of θ are as follows:

$$p_j^{(t+1)}=\frac{\sum_{i=1}^{n} T_{ji}^{(t)}}{\sum_{i=1}^{n}\left(T_{1i}^{(t)}+T_{2i}^{(t)}\right)}=\frac{1}{n}\sum_{i=1}^{n} T_{ji}^{(t)}, \qquad j=1,2 \qquad (5.3.5)$$

$$\mu_1^{(t+1)}=\frac{\sum_{i=1}^{n} T_{1i}^{(t)} x_i}{\sum_{i=1}^{n} T_{1i}^{(t)}} \quad \text{and} \quad \sigma_1^{(t+1)}=\frac{\sum_{i=1}^{n} T_{1i}^{(t)}\left(x_i-\mu_1^{(t)}\right)}{\sum_{i=1}^{n} T_{1i}^{(t)}} \qquad (5.3.6)$$

and

$$\mu_2^{(t+1)} = \frac{\sum_{i=1}^{n} T_{2i}^{(t)} x_i}{\sum_{i=1}^{n} T_{2i}^{(t)}} \quad \text{and} \quad \sigma_2^{(t+1)} = \frac{\sum_{i=1}^{n} T_{2i}^{(t)} \left(x_i - \mu_2^{(t)}\right)}{\sum_{i=1}^{n} T_{2i}^{(t)}} \tag{5.3.7}$$

5.3.6.2 Numerical Example:

We have generated the 20 observations from $N(4,9)$ and rest 20 observations from $N(20,9)$. We arranged all 40 observations in ascending order and then applied usual method and EM algorithm to estimate MLE of different parameters belonging to $\theta=(p,\mu_1,\mu_2,\sigma_1,\sigma_2)$.

The proportions for inliers are considered as 0.2 and 0.8 taking other random samples. Random numbers for inliers and target are generated from $N(10,9)$ and $N(16,9)$ for *p* = 0.2 whereas for *p* = 0.8 random numbers are generated from $N(20,9)$ and the estimates for the same are presented in the table (5.3.1).

Table 5.3.1. Estimates of parameters using usual method and EM algorithm

Parameter	Usual	EM	Usual	EM	Usual	EM
$\hat{p}$	0.20	0.247891	0.5	0.51753	0.8	0.801655
$\hat{\mu}_1$	10.3983	10.0831	5.114591	5.07216	20.7588	20.7477
$\hat{\sigma}_1$	2.1596	1.72969	2.542373	2.45262	2.04267	1.77783
$\hat{\mu}_2$	16.7081	16.4453	19.93911	19.8436	30.3582	30.3412
$\hat{\sigma}_2$	0.66221	0.74314	3.500136	3.53658	2.3719	2.32364

From above table the estimate of number of inliers $\hat{r}=n(1-\hat{p})$. We observe that estimates of usual method to obtain MLE and EM algorithm are very close for all values of *p*.

5.3.7 Newton Raphson (NR)

The purpose of NR method is same as that of EM method. This method usually requires less iteration than EM method. For NR the iterative step can be written as

$$\xi^{(m+1)} = \xi^{(m)} - \alpha_m \left[D^2 L\left(\xi^{(m)}\right) \right]^{-1} DL\left(\xi^{(m)}\right), \qquad m = 0,1\ldots\ldots\ldots \tag{5.3.8}$$

The estimation of parameters for mixture distribution is done by Newton raphson method in all chapters 2, 3 and 4 of this thesis.

5.3.8 Method of Scoring (MS)

For MS the iterative steps to obtain estimates of parameters of mixture distribution can be written as

$$\xi^{(m+1)} = \xi^{(m)} + \alpha_m \left[I\left(\xi^{(m)}\right) \right]^{-1} DL\left(\xi^{(m)}\right), \qquad m = 0,1\ldots\ldots\ldots \tag{5.3.9}$$

In above two cases, the non-negative constant α_m has been introduced to provide a slight increase in generality. Usually α_m =1, $I\left(\xi^{(m)}\right)$ denotes Fisher information matrix and D and D^2 represent differentiation, once and twice, with respectively, ξ.

We now carry out some tests of hypothesis to ascertain the model validity in the presence of inliers. We now carry out some tests of hypothesis to ascertain the model validity in the presence of inliers.

5.4 Testing of hypothesis

Goodness of fit is required to test whether the proportion of inliers and target observations considered for the model really fits in the data. Various tests are discussed in following subsections.

5.4.1 Locally most powerful test

For testing the hypothesis as defined in equation (1.4.3) we have

$$H_0 : p = 1 \text{ against } H_1 : p < 1$$

Let $X_1, X_2, \ldots\ldots X_n$ be a i.i.d. random variables having mixture distribution then likelihood is

$$L(x, \phi, \theta, p) = \prod \left\{ (1-p)\, g(x_i) + pf(x_i) \right\}$$

Then LMP test critical region is given by

$$\left[\underline{x} \,|\, \frac{\partial L(x,\phi,\theta,p)}{\partial p} \,|\, H_0 \right] \le C$$

where C is such that

$$P\left\{ \left[\underline{x} \,|\, \frac{\partial L(x,\phi,\theta,p)}{\partial p} \,|\, H_0 \right] \le C \right\} = \alpha \tag{5.4.1}$$

Solving for C we get

$$\frac{\partial \ln L(x,\phi,\theta,p)}{\partial p} = \sum_{i=1}^{n} \frac{f(x_i) - g(x_i)}{(1-p)g(x_i) + pf(x_i)}$$

$$\frac{\partial \ln L(x,\phi,\theta,p \,|\, H_0)}{\partial p} = \sum_{i=1}^{n} \frac{f(x_i) - g(x_i)}{f(x_i)}$$

$$= \sum_{i=1}^{n} \left[1 - \frac{g(x_i)}{f(x_i)} \right]$$

$$= n - \sum_{i=1}^{n} \lambda(x_i)$$

on simplification we get reject H_0 if

$$\left\{ \underline{x} \mid n - \sum_{i=1}^{n} \lambda(x_i) \le C \right\} = \left\{ \underline{x} \mid \sum_{i=1}^{n} \lambda(x_i) \ge C' \right\}$$

where C' is such that

$$P\left\{ \underline{x} \mid \sum_{i=1}^{n} \lambda(x_i) \ge C' \right\} = \alpha.$$

5.4.2 Large sample test

To test $H_0 : p \ge p_0$ against $H_0 : p < p_0$ for specified p_0, the proportion of target observations. Test statistics is given by

$$Z_{cal} = \frac{\sqrt{n}\left(\hat{p} - p_0\right)}{\sqrt{p_0 q_0}}, \quad q_0 = 1 - p_0$$

and we reject H_0 if $Z_{cal} < Z_\alpha$.

5.5 Graphs representing mixture of inliers and target distributions

In figure (5.5.1) and figure (5.5.2) we represent the graphs of density function and survival functions of mixture of two exponential distributions respectively. Here the target and inliers distribution both are exponential distribution.

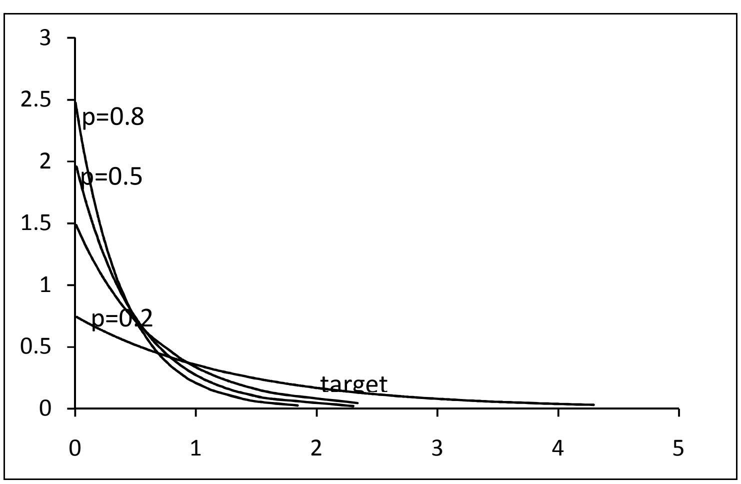


Fig. 5.5.1. Density function of exponential inliers and target distribution

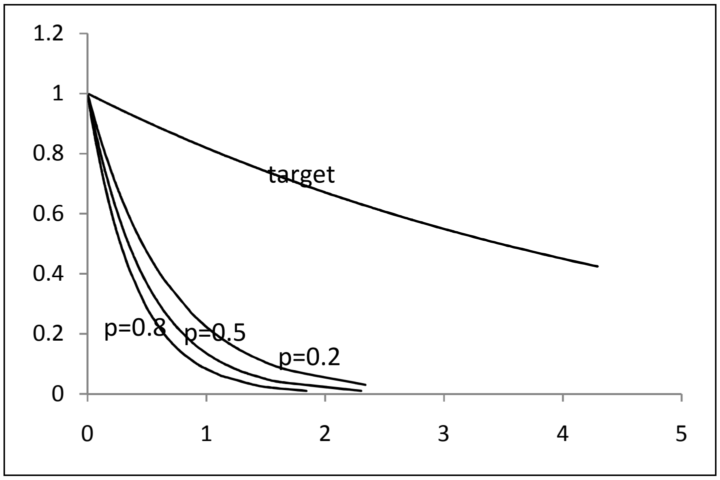


Fig. 5.5.2. Survival function of inliers and target in exponential distribution

For both the graphs we generated random sample from inliers and target population in different proportion. Here p = 0.2, 0.5 and 0.8 represents the proportion of sample from target population.

In figure (5.5.3) and (5.5.4) we have considered a random sample from a single exponential population with mean 10. Then we arranged these observations in ascending order of the magnitude. Hence we divided the observations in two parts i.e. inliers and target in different proportion and obtained the following graphs (5.5.3), (5.5.4) and estimates of the parameters with their confidence intervals in table (5.5.1).

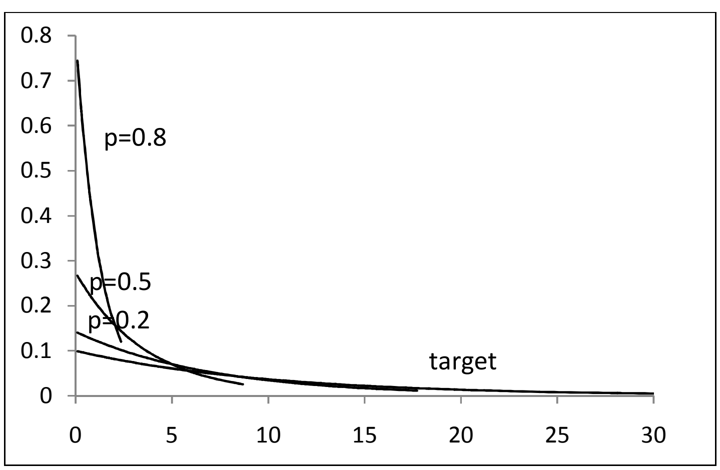


Fig. 5.5.3. Density function of exponential inliers and target distribution

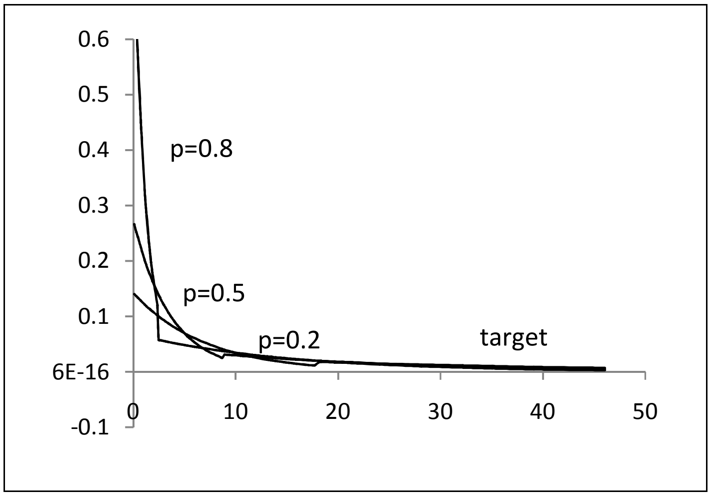


Fig. 5.5.4. Survival function of exponential inliers and target distribution

Table 5.5.1. Estimates of the parameters

p	$\hat{\phi}$	$\hat{\theta}$	confidence interval of $\hat{\phi}$	confidence interval of $\hat{\theta}$
0.2	1.23823	14.67034	(1.123822, 1.287276)	(13.95386, 15.38681)
0.5	3.649312	20.77614	(3.569648, 3.709192)	(20.54426, 21.01385)
0.8	7.014626	31.04246	(6.908973, 7.096944)	(30.89754, 31.62216)

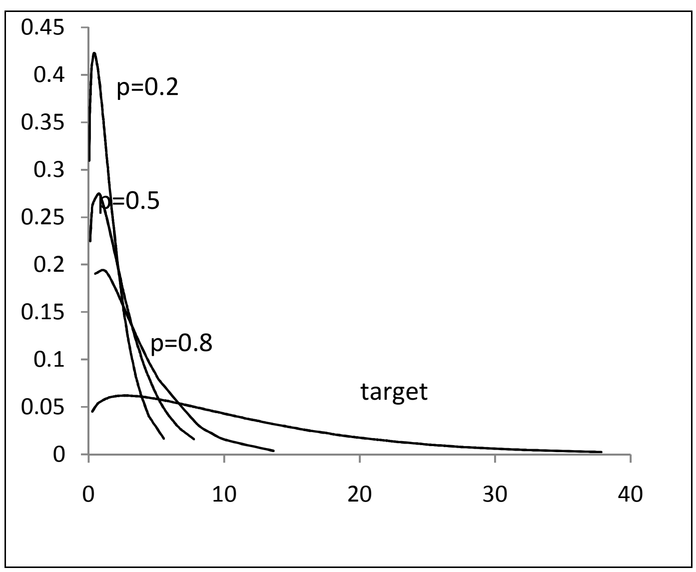


Fig. 5.5.5. Density function of Weibull inliers and target distribution

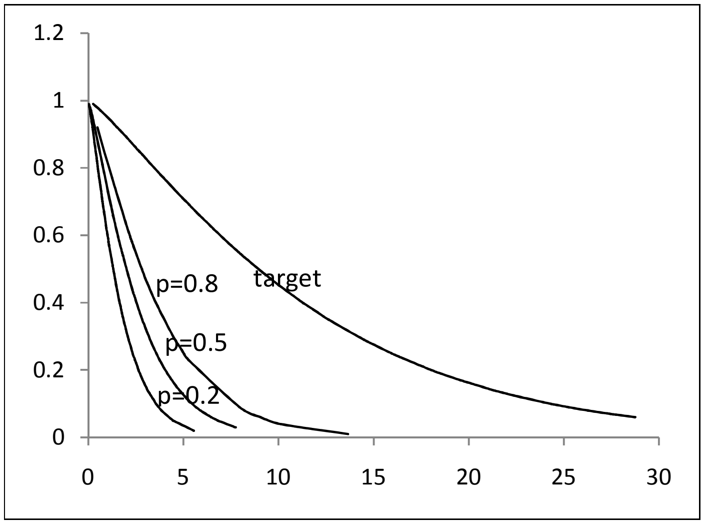


Fig. 5.5.6. Survival function of Weibull inliers and target distribution

In figure (5.5.5) and (5.5.6) we have considered generated random samples of inliers and target population from Weibull distribution. It is clear from survival function graph (5.5.2) and (5.5.6) that exponential graph has greater survival rate for target population than Weibull. Survival rate decreases more rapidly in case of Weibull distribution.

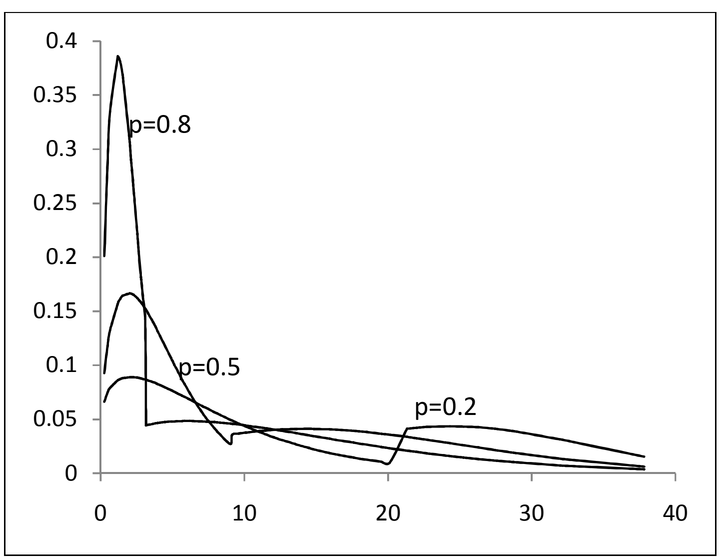


Fig. 5.5.7. Density function of Weibull inliers and target distribution

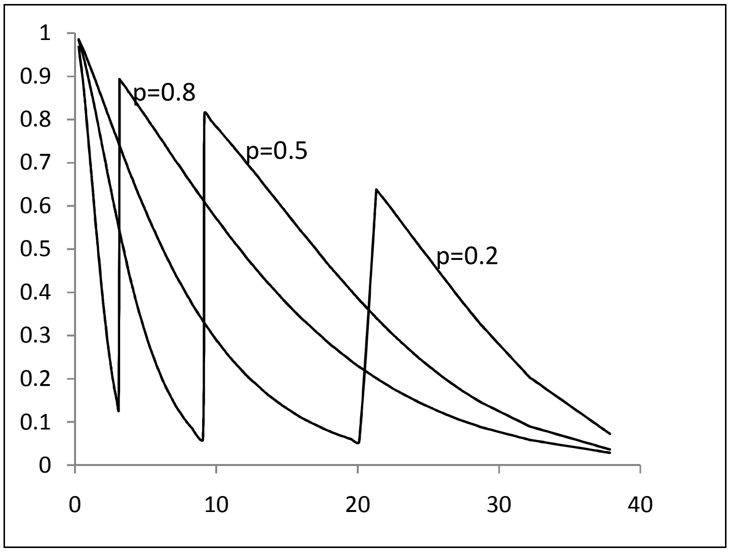


Fig. 5.5.8. Survival function of Weibull inliers and target distribution

Table 5.5.2. Estimates of the parameters(Weibull distribution)

p	$\hat{\beta}_0$	$\hat{\phi}$	$\hat{\beta}_1$	$\hat{\theta}$
0.2	1.22271	13.4815	3.07561	27153.6
0.5	1.43788	8.49143	1.95866	372.093
0.8	1.68314	3.25269	1.38857	43.6578

Rayleigh distribution for inliers and exponential for target population is considered in remaining graphs. The objective was to see how the mixture of two different distribution work. For Figure [5.5.9] and [5.5.10] the random sample of different proportion of inliers with different parameters and target observations with same parameter.

For figures [5.5.11] and [5.5.12] we have drawn two samples from Rayleigh i.e Weibull(1,2) and exponential distribution i.e exp(1). Then we took all the observation together and divided in two parts inliers and target values. From these we estimated the parameters ϕ, θ and their confidence interval for *p* = 0.0, 0.5 and 1.0.

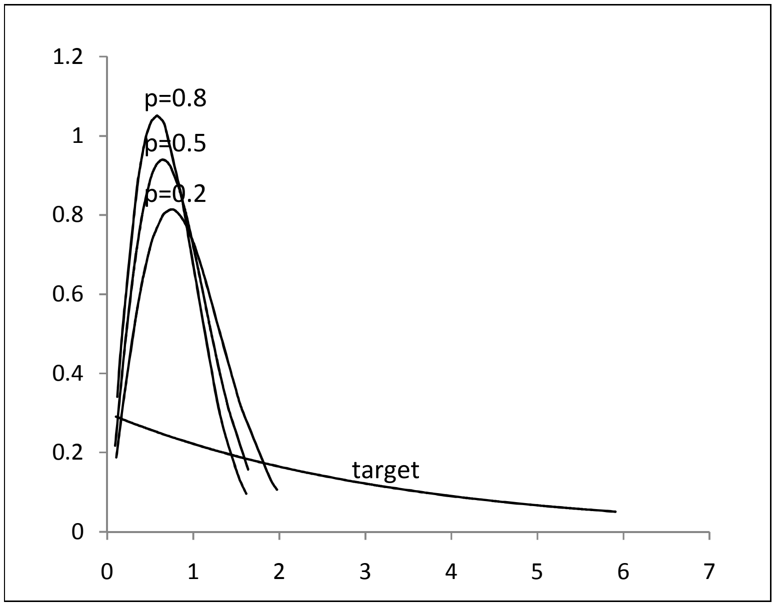


Fig. 5.5.9. Density function of Rayleigh inliers and exponential target population

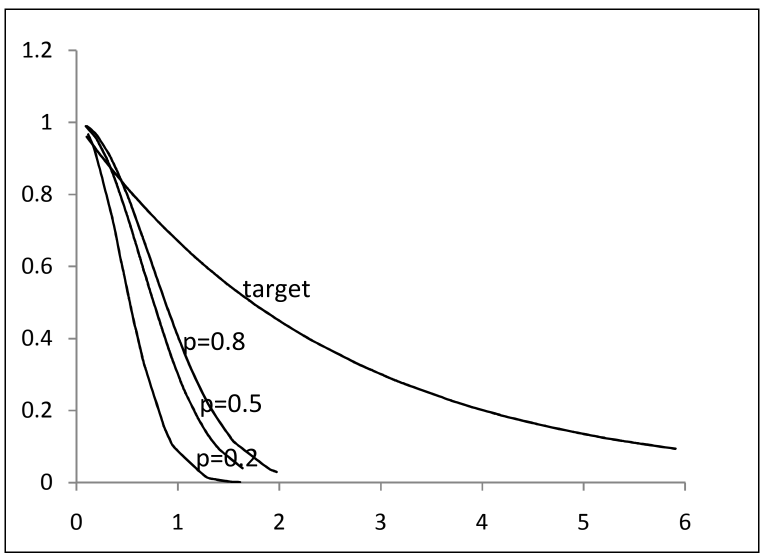


Fig. 5.5.10. Survival function of Rayleigh inliers and exponential target population

The estimates are

Table 5.5.3. Estimates of parameters

p	$\hat{\phi}$	$\hat{\theta}$	Confidence interval for ϕ	Confidence interval for θ
0.0	1.673036	____	(1.645404, 1.700668)	________
0.5	0.167334	1.479144	(0.155571, 0.179097)	(1.416527,1.541762)
1.0	______	0.973552	_________	(0.94592,0.94592)

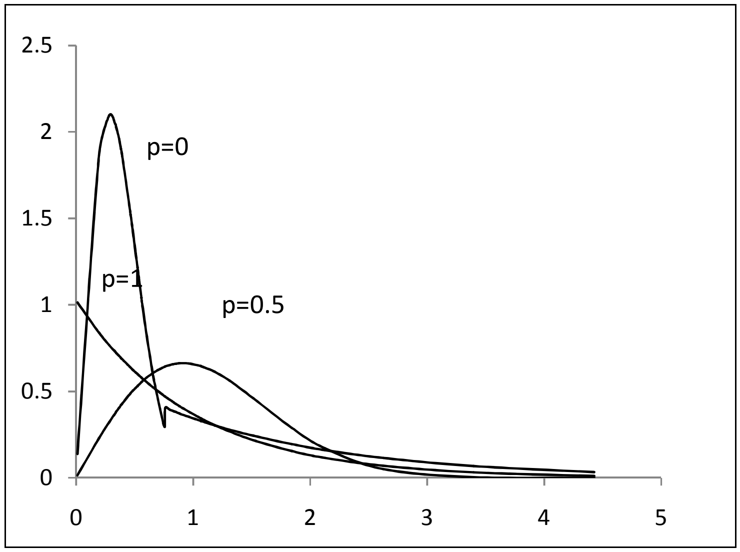


Fig. 5.5.11. Density function of Rayleigh inliers and exponential target population

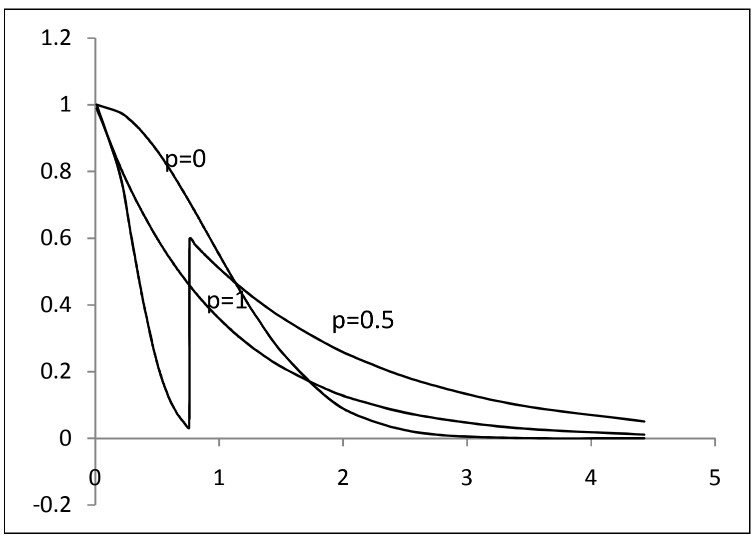


Fig. 5.5.12. Survival function of Rayleigh inliers and exponential target

Conclusion:

The discussion of mixture of two same distribution with different parameters has been studied extensively. One can also think of mixtures of two totally different distributions for inliers and target population. For example the combination of Pareto-Weibull or Normal – exponential etc., for inliers and target population, respectively.

Chapter-6

Inliers estimation in generalized failure distributions

6.1 Introduction

Generalized distributions are not frequently used for modeling life data as the life testing distribution but they have the ability to mimic the attributes of other distributions such as the exponential, Weibull or lognormal, based on the values of the distribution's parameters. Generalized exponential distribution has a right skewed unimodal density function and monotone hazard function similar to the density functions and hazard functions of the gamma and Weibull distributions. It is observed that the bivariate generalized exponential distribution provides a better fit than the bivariate exponential distribution. While the generalized gamma distribution is not often used to model life data by itself, its ability to behave like other more commonly-used life distributions is sometimes used to determine which of those life distributions should be used to model a particular set of data. It is observed that it can be used quite effectively to analyze lifetime data in place of gamma, Weibull and log-normal distributions. The genesis of this model is different estimation procedures and their properties, estimation of the stress-strength parameter, closeness of this distribution to some of the well known distribution functions, etc will be studied in this chapter from the inlier observations perspective.

6.2 Instantaneous Failures

As usual to accommodate the possibility of instantaneous failures, the class of generalized failure time distribution (GFTD) $\Im = \{F(x,\theta), \theta \in \Omega\}$ is modified to a new distribution $\mathcal{G} = \{G(x,\theta,p) = (1-p) + pF(x,\theta), F \in \Im, x \geq 0, 0 < p < 1\}$ where $f(x,\theta)$ is of the form

$$f(x,\theta,\beta) = \left(\frac{\phi(x)}{\theta}\right)^{\beta} \left[\frac{\phi'(x)}{\phi(x)}\right] \frac{1}{\overline{|\beta}} \exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right), \quad \phi(x) > 0, \theta, \beta > 0 \quad (6.2.1)$$

One may refer to Johnson and Kotz, Johnson and Balakrishnan (1970) etc. for other version of generalized densities. The above density is studied by Chaturvedi and Usha (2008).

6.2.1 Maximum likelihood estimation in instantaneous failures

The modified general failure time density function is given as

$$g(x,p,\theta,\beta) = \begin{cases} 1-p, & \phi(x) = 0 \\ \dfrac{p}{\overline{|\beta}} \left(\dfrac{\phi(x)}{\theta}\right)^{\beta} \left[\dfrac{\phi'(x)}{\phi(x)}\right] \exp\left(-\left[\dfrac{\phi(x)}{\theta}\right]\right), & \phi(x) > 0 \end{cases} \quad (6.2.2)$$

Let $X_1, X_2, ...X_n$, be a random sample of size *n* from $g \in \mathcal{G}$.

$$L(x,p,\theta,\beta) = \prod_{i=1}^{n} g(x_i, p, \theta, \beta)$$

Define

$$Z(x) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases}$$

Then the likelihood is given by

$$L(x;p,\theta,\beta)=\prod_{i=1}^{n}(1-p)^{z(x_i)}\left[pf(x_i,\theta,\beta)\right]^{1-z(x_i)}$$

$$=(1-p)^{\sum z(x_i)}\left(\frac{p}{\theta^{\beta}\overline{|\beta}}\right)^{n-\sum z(x_i)}\prod_{x_i>0}\left[\left\{\phi(x)\right\}^{\beta}\left(\frac{\phi'(x)}{\phi(x)}\right)\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right)\right]^{1-z(x_i)}$$

It is possible to show that (6.2.2) is a member of three parameter exponential family with $\left(\sum_{i=1}^{n}z(x_i),\sum\left[1-z(x_i)\right]\ln\left\{\phi(x_i)\right\},\sum\left[1-z(x_i)\right]\left\{\phi(x_i)\right\}\right)$ are jointly complete sufficient for (p,β,θ), provided $\phi(x)$ is real valued and strictly increasing function of *x* with $\phi(0)=0$ and its inverse function exists.

The estimating equations are constructed from the log likelihood and are given by

$$\frac{\partial \ln L}{\partial p}=-\frac{\sum z(x_i)}{1-p}+\frac{n-\sum z(x_i)}{p}=0 \qquad (6.2.3)$$

$$\frac{\partial \ln L}{\partial \beta}=-\left[n-\sum z(x_i)\right]\left(\frac{\partial \ln\left(\overline{|\beta}\right)}{\partial \beta}+\ln\theta\right)+\sum\left[1-z(x_i)\right]\ln\left\{\phi(x_i)\right\}=0 \qquad (6.2.4)$$

$$\frac{\partial \ln L}{\partial \theta}=\frac{\beta}{\theta}\sum z(x_i)-\frac{n\beta}{\theta}+\frac{1}{\theta^2}\sum\left[1-z(x_i)\right]\phi(x_i)=0 \qquad (6.2.5)$$

Since the equation (6.2.3) is independent of θ and β, one can solve and get $\hat{p}=\frac{n-r}{n}$, if $\sum z(x_i)=r$. The estimates $\hat{\theta}$ and $\hat{\beta}$ are obtained by solving (6.2.4) and (6.2.5) which are the conditional likelihood equations given (*n*-*r*) positive observations. One can also obtain the Fisher information as the expectation of second derivative of the likelihood equations above once the form of $\phi(x)$ is known.

6.3 Early failures

To accommodate the possibility of instantaneous and early failures the class of generalized failure time distribution (GFTD) $\Im = \{F(x,\theta), \theta \in \Omega\}$ is modified to distribution $\mathcal{G}_1 = \{G_1(x,\theta,p) = (1-p) + pF(\delta,\theta) + pf(x,\theta), F \in \Im, x \geq 0, 0 < p < 1\}$. The failure time correspond to early failures which are reported as δ which is very very small and hence the modified model will be a mixture in the proportion 1-*p* and *p.* The estimation procedure for the parameters involved in the model. The modified generalized failure time distribution is given by distribution function

$$\mathcal{G}_1(x,p,\theta) = \begin{cases} 0 & x < \delta \\ 1-p+pF(\delta,\theta), & x = \delta \\ pf(x,\theta), & x > \delta \end{cases} \tag{6.3.1}$$

which can be simplified as

$$g_1(x,p,\theta) = \begin{cases} 0 & x < \delta \\ 1-p\bar{F}(\delta,\theta), & x = \delta \\ pf(x,\theta), & x > \delta \end{cases} \tag{6.3.2}$$

6.3.2 Maximum likelihood estimation in early failures

On substituting the modified general failure time distribution is given as

$$g_1(x,p,\theta,\beta) = \begin{cases} 1-p\bar{F}(\delta,\theta,\beta), & \phi(x) = \delta \\ \dfrac{p}{\overline{|\beta}}\left(\dfrac{\phi(x)}{\theta}\right)^{\beta}\left[\dfrac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\dfrac{\phi(x)}{\theta}\right]\right), & \phi(x) > \delta \end{cases} \tag{6.3.3}$$

Let $X_1, X_2, ...X_n$, be a random sample of size *n* from $g_1 \in \mathcal{G}_1$

$$L(x,p,\theta,\beta) = \prod_{i=1}^{n} g_1(x_i,p,\theta,\beta)$$

then define

$$Z(x)=\begin{cases}1, & x=\delta\\ 0, & x>\delta\end{cases}$$

$$L(x,p,\theta,\beta)=\prod_{i=1}^{n}\left(1-p\bar{F}\{\delta,\theta,\beta\}\right)^{z(x_i)}\left[pf(x_i,\theta,\beta)\right]^{1-z(x_i)}$$

$$=\left(1-p\bar{F}(\delta,\theta,\beta)\right)^{\sum z(x_i)}\left(\frac{p}{\theta^{\beta}\overline{|\beta}}\right)^{n-\sum z(x_i)}\prod_{x_i>\delta}\left[\{\phi(x)\}^{\beta}\left(\frac{\phi'(x)}{\phi(x)}\right)\exp\left(-\left[\frac{\phi(x)}{\theta}\right]\right)\right]^{1-z(x_i)}$$

Here again it is possible to show that (6.3.3) is a member of three parameter exponential family with $\left(\sum_{i=1}^{n}z(x_i),\sum\left[1-z(x_i)\right]\ln\{\phi(x_i)\},\sum\left[1-z(x_i)\right]\{\phi(x_i)\}\right)$ are jointly complete sufficient for (p,β,θ), provided $\phi(x)$ is real valued and strictly increasing function of *x* with $\phi(\delta)=0$ and its inverse function exists. The estimating equations are constructed from the log likelihood and are given by

$$\frac{\partial\ln L}{\partial p}=-\frac{\sum z(x_i)\bar{F}(\delta,\theta,\beta)}{(1-p)\bar{F}(\delta,\theta,\beta)}+\frac{n-\sum z(x_i)}{p}=0 \tag{6.3.4}$$

$$\frac{\partial\ln L}{\partial\beta}=-\frac{p\sum z(x_i)}{(1-p)\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\beta}-\left[n-\sum z(x_i)\right]\frac{\partial\ln\left(\overline{|\beta}\right)}{\partial\beta}$$
$$+\ln\theta+\sum\left[1-z(x_i)\right]\ln\{\phi(x_i)\}=0 \tag{6.3.5}$$

$$\frac{\partial\ln L}{\partial\theta}=-\frac{p\sum z(x_i)}{(1-p)\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\theta}+\frac{\beta}{\theta}\sum z(x_i)-\frac{n\beta}{\theta}$$
$$+\frac{1}{\theta^2}\sum\left[1-z(x_i)\right]\phi(x_i)=0 \tag{6.3.6}$$

Solving (6.3.4) one gets

$$\hat{p}=\frac{(n-r)}{n\bar{F}(\delta,\theta,\beta)} \tag{6.3.7}$$

Equation (6.3.5) and (6.3.6) have to be solved simultaneously using numerical iterative method to obtain the estimation of parameters under study .

6.4 Nearly instantaneous failure

Let $F(x)$ and $R(x) = 1 - F(x)$ denote the cumulative distribution function and the survival function of the mixture, respectively. *F* is continuous and its density be given by $f(x) = F'(x)$.The component distribution functions and their survival functions are $F_i(x)$ and $R_i(x) = 1 - F_i(x)$ respectively, i=1,2. The hazard rate of a lifetime distribution is defined as $h(x) = f(x) / R(x)$ provided the density exists. Instead of assuming an instant or an early failures to occur at a particular point, as in the original model as above, we now represent this model as a mixture of the generalized Dirac delta function and the generalized failure time. Thus the resulting modification gives rise to a density function:

$$f(x) = (1-p)\delta_d(x - x_0) + \frac{p}{\lceil\beta}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{x}\right]\right) \quad ,$$

where

$$p + q = 1,\ 0 < p < 1,\ \phi(x) > 0,\ \beta > 0. \qquad (6.4.1)$$

and

$$\delta_d(x - x_0) = \begin{cases} \dfrac{1}{d}, & x_0 \le x < x_0 + d \\ 0, & otherwise \end{cases} \quad , \qquad (6.4.2)$$

for sufficiently small *d*. Here $p > 0$ is the mixing proportion. Also note that

$$\delta(x - x_0) = \lim_{d \to 0} \delta_d(x - x_0) \qquad (6.4.3)$$

where $\delta(\bullet)$ is the Dirac delta function as given in section (2.4) of chapter 2. Both the distribution and survival functions are continuous.

Writing

$$f_1(x) = \delta_d(x - x_0) \text{ and } f_2(x) = \frac{1}{\overline{|\beta}}\left(\frac{\phi(x)}{\theta}\right)^{\beta}\left[\frac{\phi'(x)}{\phi(x)}\right]\exp\left(-\left[\frac{\phi(x)}{x}\right]\right)$$

Then (6.4.1) can be written as

$$f(x) = q\,f_1(x) + p\,f_2(x) \quad \text{where} \quad p + q = 1,\ 0 < p < 1 \tag{6.4.4}$$

so that

$$F(x) = q\,F_1(x) + pF_2(x) \tag{6.4.5}$$

the corresponding survival function is

$$R(x) = 1 - F(x) = q + p - qF_1(x) - p\,F_2(x) = qR_1(x) + pR_2(x) \tag{6.4.6}$$

and the hazard function of the mixture distribution is

$$h(x) = \frac{qf_1(x) + pf_2(x)}{qR_1(x) + pR_2(x)} \tag{6.4.7}$$

Now using above results, in terms of density function of particular distribution, given in equation (6.2.2) one can obtain various characteristics.

6.4.1 Characteristics of the model

The life time models are generally characterised in terms of its hazard rate function, survival function and the mean residual life functions. Below we obtain these characteristics and obtain some useful relationship between them. The reliability (survival) functions of the respective component distributions are given by

$$R_1(x) = \begin{cases} 1\,, & 0 \le x < x_0 \\ \dfrac{d + x_0 - x}{d}, & x_0 \le x \le x_0 + d \\ undefined\,, & t \ge t_0 + d \end{cases} \tag{6.4.8}$$

and

$$R_2(x) = \overline{F}_2(x) \tag{6.4.9}$$

The hazard rates are, respectively,

$$h_1(x)=\begin{cases}0, & 0\le x<x_0\\ \dfrac{1}{d+x_0-x}, & x_0\le x\le x_0+d\\ \infty, & x\ge x_0+d\end{cases} \qquad (6.4.10)$$

and

$$h_2(x)=\frac{f_2(x)}{\bar{F}_2(x)} \qquad (6.4.11)$$

It can be shown (6.4.10) and (6.4.11) that for any mixture of two continuous distributions the hazard rate function can be expressed as

$$h(x) = \frac{f(x)}{R(x)} = w(x)h_1(x)+[1-w(x)]h_2(x) \qquad (6.4.12)$$

where $w(x)=q\,R_1(x)/R(x)$ for all $x\ge 0$. In our case,

$$w(x) = \begin{cases}\dfrac{q}{R(t)}, & 0\le x<x_0\\ \dfrac{q\,R_1(x)}{R(x)}, & x_0\le x\le x_0+d\\ 0, & x\ge x_0+d\end{cases} \qquad (6.4.13)$$

Establishing some interesting relationship between the survival function and hazard function through $w(x)$ as follows:

Since

$$w(x)=q\,R_1(x)/R(x)$$

$$w'(x)=\frac{q\left[R_1'(x)R(x)-R_1(x)R'(x)\right]}{\left[R(x)\right]^2}$$

upon substituting the value of $R(x)$ from above and simplifying, we get

$$w'(x)=\frac{pq\left[R_1'(x)R_2(x)-R_1(x)R_2'(x)\right]}{\left[R(x)\right]^2}$$

If the terms, are rearranged one gets

$$w'(x)=\frac{pqR_1(x)R_2(x)\left[\frac{R_1'(x)}{R_1(x)}-\frac{R_2'(x)}{R_2(x)}\right]}{\left[R(x)\right]^2}$$

Now recall,

$$w(x)=\frac{qR_1(x)}{R(x)},1-w(x)=\frac{pR_2(x)}{R(x)},h_1(x)=-\frac{R_1'(x)}{R_1(x)} \text{ and } h_2(x)=-\frac{R_2'(x)}{R_2(x)}$$

hence

$$w'(x)=w(x)\left[1-w(x)\right]\left\{h_2(x)-h_1(x)\right\} \tag{6.4.14}$$

in a similar way, one can show that

$$h'(x)=w'(x)h_1(x)+w(x)h_1'(x)-w'(x)h_2(x)+\left[1-w(x)\right]h_2'(x) \tag{6.4.15}$$

also, since $f_1(x)=-R_1(x)$, one gets

$$w(x)\,h_1(x)=q\frac{R_1(x)}{R(x)}\frac{f_1(x)}{R_1(x)}=\frac{qf_1(x)}{R(x)}$$

which shows that, (6.4.12) is well defined for all $x > 0$. Thus the summarized expression for $R(x)$, $h(x)$ and $m(x)$, are respectively, given as

$$R(x)=\begin{cases} q+p\bar{F}(x), & 0<x<x_0 \\ \frac{q[d+x_0-x]}{d}+p\bar{F}(x), & x_0\le x\le x_0+d \\ p\bar{F}(x), & x>x_0+d \end{cases} \tag{6.4.16}$$

$$h(x)=\begin{cases}\left[\dfrac{p\bar{F}(x)f_2(x)}{q+p\bar{F}(x)}\right], & 0\le x\le x_0\\ \dfrac{q+dp\bar{F}(x)f_2(x)}{q(d-x)+dp\bar{F}(x)}, & x_0\le x\le x_0+d\\ f_2(x), & x>x_0+d\end{cases} \tag{6.4.17}$$

The mean residual life (MRL) of a random variable X defined for all *x* as

$$m_X(x)=E(X-x/X>x)=\frac{\int_x^{\infty} R_X(y)dy}{R_X(x)}$$

This is the expected additional time to failure given survival to *x*, which can also be expressed in terms of mixture of two MRL's as

$$m(x)=qm_1(x)+pm_2(x) \tag{6.4.18}$$

where

$$m_1(x)=\begin{cases}\dfrac{x_0-x}{2}, & 0\le x<x_0\\ \dfrac{x_0+d-x}{2}, & x_0\le x<x_0+d\\ 0, & x>x_0+d\end{cases} \tag{6.4.19}$$

and

$$m_2(x)=\frac{\int_x^{\infty}\bar{F}_2(y)dy}{\bar{F}_2(x)}, \qquad y>d \tag{6.4.20}$$

6.4.2 Particular Case When (X_0 = 0)

Consider a special case of model (6.4.1) whereby x_0 = 0. The model may be called the model with "nearly instantaneous failure". In this case, (6.4.3) is simplified giving the hazard rate of the uniform distribution as

$$h_1(x)=\begin{cases}\frac{1}{d-x}, & 0\le x\le d\\ \infty, & x>d\end{cases} \tag{6.4.21}$$

and its survival rate function is given as

$$R_1(x)=\begin{cases}\frac{d-x}{d}, & 0\le x\le d\\ 0, & x>d\end{cases} \tag{6.4.22}$$

Thus the generalized model with "nearly instantaneous failure" occurring uniformly over [0, *d*] has the survival function

$$R(x)=\begin{cases}\frac{q(d-x)}{d}+p\bar{F}_2(x), & 0\le x\le d\\ p\bar{F}_2(x), & x>d\end{cases} \tag{6.4.23}$$

and the hazard function as

$$h(x)=\begin{cases}\frac{q+dpf_2(x)}{q(d-x)+dp}, & 0\le x\le x_0+d\\ f_2(x), & x>x_0+d\end{cases} \tag{6.4.24}$$

One can study the above characteristics by plotting graphs, with various combinations of values of parameters.

6.5 Testing of hypothesis

Here the interest is to test the hypothesis, whether sample observations belong to inliers population against hypothesis that it belongs to target population. Refer equation (1.6.1), the hypothesis can be written as

$$H_0:\xi=\phi \text{ versus } H_0:\xi\neq\phi. \tag{6.5.1}$$

where ξ is the common population parameter under study. Below we discuss various computationally simple test procedures to detect inliers in a model.

6.5.1 Sequential Probability Ratio Test (SPRT) to detect inliers in the model

SPRT to test the hypothesis whether a observation belongs to inlier population with p.d.f. $g(x,\phi)$ against hypothesis that it belongs to target population with p.d.f. $f(x,\theta)$. i.e. equation (6.5.1).

The likelihood when H_1 is true , is given by

$$L_1 = \prod_{i=1}^{r} f(x_i,\theta)$$

and under H_0, it is

$$L_0 = \prod_{i=1}^{r} g(x_i,\phi)$$

And likelihood ratio λ_r is given by $\lambda_r = \frac{L_1}{L_0}$ or equivalently

$$\ln\lambda_r = \sum_{i=1}^{r} \ln\frac{f(x_i,\theta)}{g(x_i,\phi)} = \sum_{i=1}^{r} z_i \qquad (6.5.2)$$

For deciding number of inliers we continue to take ordered observations one by one till we reject H_0. That is

if $\sum_{i=1}^{r} z_i \le \log B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{r} z_i \ge \log A$ reject H_0 and stop.

The corresponding m represents the first observation from target observation and hence $\hat{r} = r-1$ are the number of inliers. And $B = \frac{\gamma}{1-\alpha}$, $A = \frac{1-\gamma}{\alpha}$, where α represents probability of type I error and γ represents probability of type II error. Now the SPRT procedure is investigated for following special cases.

Case 1: Testing for scale parameter when shape parameter $\beta_0 = \beta_1 = b$.

To test: $H_0:\theta=\theta_0$ against $H_1:\theta=\theta_1$

θ_0 and θ_1 are the scale parameters of inlier distribution and target distribution respectively.

The test statistics is

$$\ln\lambda_{r\theta}=r\beta_0\left[\ln\theta_0-\ln\theta_1\right]+\sum_{i=1}^{r}\phi(x)\left[\frac{1}{\theta_0}-\frac{1}{\theta_1}\right] \tag{6.5.3}$$

Reject H_0 when

$$\sum_{i=1}^{r}\phi(x)>\frac{\ln A-r\beta_0\left[\ln\theta_0-\ln\theta_1\right]}{\left[\frac{1}{\theta_0}-\frac{1}{\theta_1}\right]} \tag{6.5.4}$$

Case 2: Testing for shape parameter when scale parameter $\theta_0=\theta_1=\theta$.

To test : $H_0:\beta=\beta_0$ against $H_1:\beta=\beta_1$

β_0 and β_1 are the shape parameters of inliers distribution and target distributions respectively.

The test statistics is

$$\ln\lambda_{r\beta}=\left[r\beta_0-r\beta_1\right]\left\{\ln\theta-\sum_{i=1}^{r}\ln\phi(x)\right\}+r\left[\ln\overline{\beta_0}-\ln\overline{\beta_1}\right] \tag{6.5.5}$$

Reject H_0 when

$$\sum_{i=1}^{r}\ln\phi(x)>\frac{\ln A-r\left[\ln\overline{\beta_0}-\ln\overline{\beta_1}\right]}{r\left[\beta_1-\beta_0\right]}+\ln\theta \tag{6.5.6}$$

6.5.2 Most Powerful Test

For the hypothesis as defined in equation (6.5.1), the most powerful test to reject H_0 is given by

$$\psi(x)=\begin{cases}1, & \dfrac{P_1(x)}{P_0(x)}>C\alpha\\[2ex] 0, & \dfrac{P_1(x)}{P_0(x)}<C\alpha\end{cases} \tag{6.5.7}$$

Where $P_1(x)$ and $P_0(x)$ are likelihood functions under distribution of target population $\Im$ and inlier population $\mathcal{G}$ respectively $C\alpha$ is such that test attains level of the test when H_0 is true. We reject H_0 for large values of the ratio $\dfrac{P_1(x)}{P_0(x)}$.

Case 1: Testing for scale parameter when shape parameter $\beta_0=\beta_1=b$.

To test $H_0:\theta=\theta_0$ against $H_1:\theta=\theta_1$ where the parameters are as defined before in section (6.5.1).

The most powerful test is given by

$$\psi(x)=\begin{cases}1, & \sum_{i=1}^{n}\phi(x)>\dfrac{C_\alpha-n\beta[\log\theta_0-\log\theta_1]}{\left[\dfrac{1}{\theta_0}-\dfrac{1}{\theta_1}\right]}\\[2ex] 0, & \textit{otherwise.}\end{cases} \tag{6.5.8}$$

Case 2: Testing for shape parameter when scale parameter $\theta_0=\theta_1=\theta$.

To test : $H_0:\beta=\beta_0$ against $H_1:\beta=\beta_1$

The most powerful test is given by

$$\psi(x)=\begin{cases}1, & \sum_{i=1}^{n}\log\phi(x)>\dfrac{C_\alpha-n\left[\log\overline{\beta_0}-\log\overline{\beta_1}\right]}{n[\beta_1-\beta_0]}+\log\theta\\[2ex] 0, & \textit{otherwise.}\end{cases} \tag{6.5.9}$$

$C\alpha$ is such that test attains level of the test when H_0 is true. Once $C\alpha$ is obtained we can find power function under alternative hypothesis.

6.6 Information Criterion

Again three information criteria which are already discussed in section (2.5) are used such as Schwarz's Information criterion (*SIC= -2lnL(Θ) + p ln n),* Schawarz's Bayesian Information criterion (*BIC= -lnL(Θ) + 0.5(p ln n)/n*) and Hannan-Quinn criterion (*HQ= -lnL(Θ) + p ln[ln(n)])* to detect number of inliers. *L*(*Θ*) represents the maximum likelihood function and *p* is the number of free parameters that need to be estimated under the model.

Below we develop the procedure for *SIC* scheme. We consider the model of no inliers as Model *SIC*(0), where all the observations are from target population. Model *SIC*(*r*) will denote *r* observations are inlier and remaining (*n-r*) observations are from target population. Our aim is to obtain number of inliers in the sample. For density in equation (6.2.2) the model with zero inlier is given by

$$SIC(0)=2n\beta_1\ln\theta_1-2(\beta_1-1)\sum_{i=1}^{n}\ln\phi(x_i)+2n\ln\Gamma\beta_1+\frac{2\sum_{i=1}^{n}\phi(x_i)}{\theta_1}+2\ln n \qquad (6.6.1)$$

and the model with *r* inliers is as

$$SIC(r)=2\beta_0 r\ln\theta_0-2(\beta_0-1)\sum_{i=1}^{r}\ln\phi(x_i)+2n\ln\Gamma\beta_0+\frac{2\sum_{i=1}^{r}\phi(x_i)}{\theta_1}+2\beta_1(n-r)\ln\theta_1$$

$$-2(\beta_1-1)\sum_{i=r+1}^{n}\ln\phi(x_i)+2(n-r)\ln\Gamma\beta_1+\frac{2\sum_{i=r+1}^{n}\phi(x_i)}{\theta_1}+4\ln n \qquad (6.6.2)$$

According to the procedure, the model(0) is selected with no inliers if $SIC(0)<\min_{1\le r\le n-1}SIC(r)$. And the model(r) is selected if $SIC(0)>\min_{1\le r\le n-1}SIC(r)$. Similarly we can find criteria for *BIC* and *HQ*.

6.7 Estimation and test for specific distributions

One can obtain, life distribution, such as, exponential, gamma, Weibull and Rayleigh distribution by substituting appropriate form of the parameters.

6.7.1 Exponential model

If $\phi(x)=x$ and $\beta=1$ then (6.2.2) reduces to a one parameter exponential distribution with the MLE under instantaneous failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i}{n-r}. \tag{6.7.1}$$

The MLE under early failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{ne^{-\frac{x}{\theta}}}, \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>\delta} x_i}{n-r}-\delta. \tag{6.7.2}$$

The test criteria to test H_0 : There are no inliers in data set against a single inlier is present in the data from exponential distribution is given by

Reject H_0

$$\text{if} \qquad \frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}}<C \tag{6.7.3}$$

where C is to be chosen such that

$$P\left(\frac{x_{(1)}}{\sum_{i=1}^{n} x_{(i)}}<C\right)=\alpha,$$

where α is the size of the test.

6.7.2 Rayleigh model

If $\phi(x)=x^2$ and $\beta=1$ then (6.2.2) reduces to a one parameter Rayleigh distribution , and the MLE under instantaneous failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^2}{n-r}. \tag{6.7.4}$$

The MLE under early failure model for $\hat{p}$ and $\hat{\theta}$ is given as

$$\hat{p}=\frac{n-r}{n}e^{\frac{x^2}{\theta}} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>\delta} x_i^2}{n-r}-\delta^2 \tag{6.7.5}$$

To test H_0 : all observations are from Rayleigh distribution with parameter θ against a single inlier is present in the data, is given by

Reject H_0

$$\text{if} \qquad \frac{x_{(1)}^2}{\sum_{i=1}^{n} x_{(i)}^2}<C \tag{6.7.6}$$

where C is to be chosen such that we attain the size of the test under null hypothesis.

6.7.3 Weibull model

If $\phi(x)=x^b$ and $\beta=1$ then (6.2.2) reduces to a two parameter Weibull distribution. The MLE under instantaneous failure model for p, b and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^b}{n-r}. \tag{6.7.7}$$

and for the estimate of b one has to solve the following equation

$$\frac{(n-r)}{b}+\sum_{x_i>0}\ln x_i-\frac{(n-r)\sum_{x_i>0} x_i^b \ln x_i}{\sum_{x_i>0} x_i^b}=0 \tag{6.7.8}$$

similarly the MLE under early failure model for p and θ is given as

$$\hat{p}=\frac{n-r}{n}e^{\frac{x^b}{\theta}} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i^b}{n-r}-\delta^b \tag{6.7.9}$$

and for parameter *b* one has to solve the following equation

$$\frac{(n-r)\delta-\sum_{x_i>\delta} x_i^b \ln x_i}{\sum_{x_i>\delta} x_i^b-(n-r)\delta^b}+\frac{1}{b}+\frac{\sum_{x_i>\delta}\ln x_i}{(n-r)}=0 \tag{6.7.10}$$

The test for presence of single inlier in Weibull family is derived in section (6.7.4.1).

6.7.4 Gamma model

If $\phi(x)=x$ then (6.2.2) reduces to a two parameter Gamma distribution with the MLE under instantaneous failure model for p, β and θ is given as

$$\hat{p}=\frac{n-r}{n} \quad \text{and} \quad \hat{\theta}=\frac{\sum_{x_i>0} x_i}{(n-r)\beta} \tag{6.7.11}$$

where for β one has to solve the equation

$$\frac{(n-r)}{\beta}-(n-r)\frac{\partial \ln(\lceil\beta)}{\partial\beta}+\ln\theta+\sum_{x_i>0}\ln x_i=0 \tag{6.7.12}$$

The MLE under early failure model for *p* and θ is given as

$$\hat{p}=\frac{n-r}{n\bar{F}(\delta,\theta,\beta)} \tag{6.7.13}$$

For getting the estimates of θ,β the following two equations are obtained, which are to be solved simultaneously

$$\frac{rp}{1-p\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\beta}-(n-r)\frac{\partial\ln(\lceil\beta)}{\partial\beta}+\ln\theta+\sum_{x_i>0}\ln x_i=0 \tag{6.7.14}$$

$$\frac{rp}{1-p\bar{F}(\delta,\theta,\beta)}\frac{\partial\bar{F}(\delta,\theta,\beta)}{\partial\theta}-(n-r)\frac{\beta}{\theta}+\frac{\sum_{x_i>\delta}x_i}{\theta^2}=0 \qquad (6.7.15)$$

where

$$\bar{F}(\delta,\theta,\beta)=1-\frac{x^{\beta}}{\theta^{\beta}}\frac{\left(\Gamma\beta-\Gamma\left(\beta,\frac{x}{\theta}\right)\right)}{\Gamma\beta}.$$

The test for single inlier in Gamma distribution is equivalent to that of exponential distribution.

6.7.4.1 Testing for one inlier in Weibull family

Consider the problem of $H_0: r=0$ (i.e. no inliers) versus $H_0: r=1$ (i.e. one inlier) in data with Weibull distribution. The joint pdf under H_0 is given by L_0 and under H_1 is given by L_1. Hence

$$L_1=c\phi\theta^{n-1}\exp\left(-\left[\frac{x_{(1)}}{\phi}+\frac{\sum_{i=2}^{n}x_{(i)}}{\theta}\right]\right) \qquad (6.7.12)$$

and

$$L_0=c\theta^{n}\exp\left(-\left[\frac{\sum_{i=2}^{n}x_{(i)}}{\theta}\right]\right) \qquad (6.7.13)$$

We already know that $\hat{\theta}=\frac{\sum_{1=1}^{n}x_{(i)}}{n}$ under H_0 where as under H_1 $\hat{\theta}=\frac{\sum_{1=2}^{n}x_{(i)}}{(n-1)}$ and $\hat{\phi}=x_{(1)}$.

Substituting the above values in equations (6.7.12) and (6.7.13), one can obtain likelihood ratio test as

$$\frac{L_0}{L_1} < C \quad \Rightarrow \quad \frac{\left(\sum_{i=2}^{n} x_{(i)}^{\beta}\right)^{n-1}}{\left(\sum_{i=1}^{n} x_{(i)}^{\beta}\right)^{n}} x_{(1)}^{\beta} < C \tag{6.7.14}$$

$$= \quad \frac{\left(\sum_{i=1}^{n} x_{(i)}^{\beta} - x_{(1)}^{\beta}\right)^{n-1}}{\left(\sum_{i=1}^{n} x_{(i)}^{\beta}\right)^{n}} x_{(1)} < C$$

$$= \quad \frac{\left(T \ - x_{(1)}^{\beta}\right)^{n-1}}{\left(T\right)^{n-1}} \frac{x_{(1)}^{\beta}}{T} < C \quad \text{where} \quad T = \sum_{1=1}^{n} x_{(i)}^{\beta} .$$

The test is to reject H_0 if

$$\frac{x_{(1)}^{\beta}}{T} < C \tag{6.7.15}$$

where C is to be chosen such that

$$P\left(\frac{x_{(1)}^{\beta}}{T} < C\right) = \alpha ,$$

and α is the size of the test. For the simulated data from Weibull (0.02, 5) the values of C for various size of the test are obtained in table (6.7.1). Using these C values power of the test have been obtained in table (6.7.2). For computation of power the data is simulated from Weibull (0.001, 1).

Table 6.7.1. Values of *C*

α	n			
	10	30	50	100
0.01	0.00065	0.00023	0.00015	7.64E-05
0.025	0.00095	0.00027	0.00016	8.14E-05
0.05	0.00149	0.00029	0.00017	8.44E-05
0.10	0.00290	0.00051	0.0002	9.09E-05
0.25	0.00670	0.00089	0.00038	0.0001
0.95	0.02343	0.00324	0.00119	0.00036
0.99	0.03555	0.00499	0.00189	0.00051

Table 6.7.2. Power of the test

n			
10	30	50	100
0.004	0.005	0.007	0.008
0.036	0.035	0.046	0.047
0.083	0.073	0.095	0.098
0.197	0.234	0.242	0.245
0.423	0.467	0.492	0.494
0.888	0.934	0.945	0.947
0.978	0.991	0.989	0.99

The power in above table are found using C values obtained in table (6.7.1).

6.8 Application

The data, collected by Amutha and Porchelvan (2009), represents monthly rainfall (in mm) during year 2004 and 2006 for the estimation of surface runoff in

Malattar Sub-watershed in Andhra Pradesh. The watershed experiences tropical monsoon climate with normal temperature, humidity and evaporation throughout the year. Runoff is one of the important hydrologic variables used in water resources applications and management planning. For gauged watershed accuracy of estimation of runoff on land surface and river requires much time and effort.

Set 1 (2004) : 3.40, 0.00, 0.00, 15.80, 232.80, 8.80, 123.20, 47.00, 154.00, 103.20, 89.80 and 12.20.

Set 2 (2006) : 0.00, 0.00, 21.40, 60.20, 53.86, 93.20, 27.80, 45.40, 205.40, 101.20, 128.20 and 0.00.

Using Kolmogorov-Smirnov test, we have come to the conclusion that exponential distribution fits well to above set 1 and set 2. Hence the analysis for the data set 1 and 2 is conducted for Exponential and Rayleigh distribution. Estimates of parameters with their standard error are calculated for instantaneous failure, early failures and nearly instantaneous model shown in the tables below.

Table 6.8.1. Instantaneous Failures

Distribution	Parameter	Set 1		Set 2	
		Estimates	Standard Error	Estimates	Standard Error
	$\hat{p}$	0.83333	2.68328	0.75	0.012217
Exponential	$\hat{\theta}$	0.01265	0.00400	2.30941	0.003863
Rayleigh	$\hat{\theta}$	0.000166	5.24224E-05	0.000103011	3.43E-05

Table 6.8.2. Early Failures

Distribution	Parameter	Set 1 $\delta = 20$		Set 2 $\delta = 30$	
		Estimates	Standard Error	Estimates	Standard Error
	$\hat{p}$	0.50000	2.00000	0.58333	2.02837
Exponential	$\hat{\phi}$	0.09950	0.04062	0.04065	0.016595
	$\hat{\theta}$	0.00799	0.00325	0.010182	0.004157
Rayleigh	$\hat{\phi}$	0.008205	0.00259	0.001624	0.001149
	$\hat{\theta}$	0.00010	3.170E-05	8.126E-05	3.071E-05

Table 6.8.3. Nearly Instantaneous Failures

Distribution	Parameter	Set 1 δ=20		Set 2 δ=21.4	
		Estimates	Standard Error	Estimates	Standard Error
Exponential	$\hat{\theta}$	0.01481	0.00428	0.01583	0.00457
Rayleigh	$\hat{\theta}$	0.000104	3.3E-05	0.000136	3.95E-05

From above table we can clearly observe that Rayleigh distribution fits better to above data sets.

6.9 Future Prospects

We have considered the Bayesian approach to inliers problem only for exponential model in this chapter. Also considered in this chapter is the inlier estimation of mixture of two different distributions from exponential family. This is further extended for mixture of any two life testing distributions when inliers are encountered. Bayesian method for estimation of parameters of mixture distribution of inliers and target population, assuming distribution other than exponential is also explored. It is possible to have observations as inliers, target and outliers, thus leading to mixture of three densities. The estimation procedure for such a model is challenging. We will be pursuing this study in future.

Appendix

Example 1:

The main reason for detecting early failures is that the inclusion of these observations will result in underestimating life expectancy or the reliability of the item or system. This in turn may underestimate the true quality of the product. But there are situations in which instantaneous or early failures may be desirable. For example, consider the following experiment carried by Vannman (1991). A batch of wooden boards is dried by a particular chemical process and the object of the experiment is to compare two processes as regards the extent of deformation of boards due to checking. The measure of damage to the board is the checking area *x* defined as $x = \frac{l\bar{d}}{hl_0}100$, where *l* is the length of the check, $\bar{d}$ is the mean depth of the check, *h* is the thickness of the board area and l_0 is the length of the board. Thus *x* is the check area measured as percentage of the board area. The boards are dried at the same time under different schedule and under some climatic conditions. When drying boards not all of them will get the checks and a typical sample of wood contain several observations with x_i = 0 or x_i > 0 but relatively small compared to the rest of the checks. These observations will correspond to instantaneous failures or early failures. Note that the larger the number of instantaneous failures better is the process. Below is the reproduced data of Schedule 1 and 2 of Experiment 3 conducted by Vannman (1991). In both the case *n*=37.

E-3, S-1: $x_i = 0$, i=1,2,…,13 and the other positive observations arranged in increasing order are 0.08, 0.32, 0.38, 0.46, 0.71, 0.82, 1.15,1.23, 1.40, 3.00, 3.23, 4.03, 4.20, 5.04, 5.36, 6.12, 6.79, 7.90, 8.27, 8.62, 9.50, 10.15, 10.58 and 17.49.

E-3, S-2: $x_i = 0$, i=1,2,…,17 and the other 20 positive observations arranged in increasing order are 0.02, 0.02, 0.02, 0.04, 0.09, 0.23, 0.26, 0.37, 0.93, 0.94, 1.02, 2.23, 2.79, 3.93, 4.47, 5.12, 5.19, 5.39, 6.83 and 8.22.

Bibliography

Abramovitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions*, General Publishing Company Ltd, Canada.

Adler, R. J. (1990). *An introduction to continuity, extrema, and related topics for Gaussian processes,* Institute of Mathematical Sciences.

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin, *J. Amer. Stat. Assn*. Vol. 50, pg. 901-908.

Akaike, H. (1974). A new look at the Statistical identification model, *IEEE Trans. Auto. Control,* 19, pg. 716-723.

Akaike, H., (1977). "On entropy maximization principle". In: Krishnaiah, P.R. (Editor), *Applications of Statistics*, North-Holland, Amsterdam, pg. 27–41.

Akkaya, A. K. and Tiku, M. L. (2005). Robust Estimation and hypothesis testing under short-tailedness and inliers, *Sociedad de Estadistica e Investigacion Operativa*, Vol 14, No. 1, pg. 129-150.

Amutha, R. and Porchelvan, P. (2009). Estimation of surface runoff in Malattar Sub-watershed using SCS-CN method, *Journal Of the Indian Society of Remote sensing,* 37, pg. 291-304.

Anderson, T. G. (1996). Return of voltality and trading volume: an information flow interpretation of stochastic voltality, *Jounrnal of finance*, 51, pg. 169-204.

Arcidiacono, P. & Jones, J. B. (2003). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm, *Econometrica, Econometric Society*, vol.71(3), pg. 933-946.

Barnett, V. and Lewis, T. (1994) *Outliers in Statistical Data.* John Wiley.

Bebbington, M., Lai C. D. and Zitkis, R. (2007). Modeling human mortality using mixtures bathtub shaped failure distributions, *Journal of Theoretical Biology,* Vol.245, pg. 528-538.

Bendre, S. M. and Kale, B. K. (1987). Masking effect on test for outliers in normal samples, *Biometrica,* 74(4), pg. 891-896.

Bickel, P. J. and Doksum, K. (1969). Tests for monotone failure based on normalized spacings, *Annals of Mathematical Statistics,* 40, pg. 1216-1235.

Block, H. W., Li, Y. and Savits, T. H. (2003). Initial and final behavior of the failure rate functions for mixtures and systems, *Journal of Applied Probability,* Vol.40, pg. 721-740.

B¨ohning, D. (1999). The lower bound method in probit regression, *Computational Statistics and Data Analysis*, 30, pg. 13–17.

Box, G. E. P. (1980). Sampling and Bayesian inference in scientific modelling and robustness, *Journal of the Royal Statistical Society, A*, Vol.143, pg. 383–430.

Bowman, K. O. Shenton, C. R. (1988). *Properties of estimators for Gamma Distributions*, Marcel Dekker. Inc. New York.

Chaturvedi, A and Sharma, V. (2008). Bayesian life test planning for a family of lifetime distributions: some approximate solutions. *Journal of Probability and Statistical Science*: Accepted for Publication.

Chen, J. (1998). Penalized likelihood ratio test for finite mixture models with multinomial observations. *Canadian Journal of Statistics*, 26, pg. 583-599.

Chen, J. and Li, P. (2009). Hypothesis test for normal mixture models: The EM approach*, Annals of Statistics,* Vol. 37, No. 5A, pg. 2523-2542.

Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter, *Journal of Statistical Planning and Inference,* 129, pg. 93-107.

Chen, H. and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures*. Statistical Probability Letters,* 52, pg. 125-133.

Chen, R. and Liu, J. S. (2002). Mixture Kalman filters, *Journal of the Royal Statistical Society Series B (Statistical Methodology),* Vol 62, Issue 3, pg. 493–508.

Chernoff, H. and Lander, E. (2009). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial, *Journal of Statistical Planning and inference,* 43, pg. 19-40.

Corduneanu, A. and Christopher, M. B. (2001). Variational Bayesian Model Selection for Mixture Distributions*, Artificial Intelligence and Statistics*, pg. 27-34.

Cramer, H. (1966). On stochastic processes whose tracjectories have no discontinuities of second type, *Annals of Mathematics*, IV, 71, pg. 85-92.

Dacunha, C. D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization : population mixtures and stationary ARMA processes, *Annals of Statistics*, Vol. 27, No. 4, pg. 1178-1209.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from incomplete data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological),* Vol 39, No.1, Pg. 1-38.

Dick, N. P. & Bowden, D. C. (1973). Maximum likelihood estimation for mixtures of two normal distributions. *Biometrics,* 29, pg. 781-790.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling, *Journal of the Royal Statistical Society B Methodological,* Vol. 56, Issue 2, Pages: 363-375.

Escobar, M. D. and West, M. (1998). *Computing nonparametric hierarchical models*, *CiteseerX.*

Everitt, B. S. (1984). Maximum Likelihood Estimation of the Parameters in a Mixture of Two Univariate Normal Distributions; A Comparison of Different Algorithms, *Jour. Of the Royal Statis. Soc., D, (The Statistician)* Vol. 33, No. 2, pg. 205-215.

Everitt, B. S. and Bullmore, E. T. (1999). Mixture Model Mapping of Brain Activation in functional Magnetic Resonance Images, *Human Brain Mapping* Vol 7, pg. 1–14.

Fisher, M. J., Masi, D. M. B., Gross, D. and Shortle, J. F. (2005). One parameter Pareto, two parameter Pareto, Three parameter Pareto: Is there a modeling difference?, *The Telecommunications Review*, pg. 79-91.

Garel, B. (2001). Likelihood Ratio Test for Univariate Gaussian Mixture, *Journal of Statistical Planning and Inference (JSPI)*, 96(2), pg. 325-350.

Gare,l B. & Goussanou, F. (2002). Removing the separation condition in a 1 against 3 components Gaussian mixture problem, *Classification, Clustering and Data Analysis*, K. Jajuga, A. Sokolowski and H-H. Bock Eds., *Springer*, Vol.6, pg. 61-73.

Geisser. S, and Eddy, W. F. (1979). A predictive approach to model selection, *Journal of American Statistical Association,* 74, pg. 153-160.

Ghosh, J. and Sen, P. (1985) On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In L. LeCam and R. A. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, pg. 789-806.

Hartigan, J. A. (1985). A failure of likelihood asymptotics for normal mixtures. In L. LeCam and R. A. Olshen (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, pg.. 807-810.

Hawkins, D. (1980), *Identification of Outliers*, Chapman and Hall,.

Hussaini, A. L. (2003), Bayesian predictive density of order statistics based on finite Mixture models, *Journal of Statistical Planning and inference*. Vol. 113, Issue 1, Pg. 15-24.

Jayade, V. P. and Prasad, M. S. (1990). Estimation of parameters of mixed failure time distribution, *Comm. Statist.Theory and Methods*, 19(12), pg. 4667-4677.

Jian-Ming Mo and Zong-Fang. (2008). Optimal Selection of Loss Severity Distribution Based on LDA, *Proceedings of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management ,* Vol.2, pg. 570-574.

Johnson, N. L., Kotz S., Balakrishnan, N. (1970). *Continuous Univariate Distributions-I*, John Wiley and Sons, New York.

Kale, B.K. (1999). A *first course on Parametric inference*, Narosa Publishers, New Delhi.

Kale, B. K. and Muralidharan, K. (2000). Optimal estimating equations in mixture distributions accommodating instantaneous or early failures, *J.Indian . Statist. Assoc*., 38, pg. 317-329.

Kale, B. K. (2001). *Industrial Mathematics and Statistics*, Narosa Publishers, New Delhi.

Kale**,** B. K. and Muralidharan, K. (2007)**.** Masking effect of inliers, *J. Indian Statistical Association,* 45(1), pg. 33-49.

Kale, B. K. and Muralidharan, K. (2007). Maximum Likelihood estimation in presence of inliers, *Journal of Indian Society for Probability and Statistics*, 10, pg. 65-80.

Kleyle, R.M. and Dahiya, R.L. (1975). Estimation of parameters of mixed failure time distribution from censored data, *Comm. Statist. Theory and Methods*, 4(9), 873-882.

Kullback, S. and Leibler, R. A. (1951). On the information and sufficiency, *Annals of Mathematical Statistics*, 22, pg. 79-86.

Lai, C. D., Khoo, B. C., Muralidharan, K. and Xie, M. (2007). Weibull model allowing nearly instantaneous failures*, J. Applied Mathematics and Decision Sciences,* ID 90842, 11 pages.

Lai Li-Hua , Wu Pei-Hsuan (2008). Estimating the threshold value and loss distribution: Rice damaged by typhoons in Taiwan*, African Journal of Agricultural Research,* Vol. 3 (12), pg. 818-824.

Lee Wo-Chiang. (2009). Applying Generalized Pareto Distribution to the Risk Mgt. of Commerce Fire Insurance, Preliminary paper.

Lehmann, E. L. and Romano, J. P. (1959). *Testing Statistical Hypotheses*, Springer Sciences.

Lian, G. S. (2009). Parametric and nonparametric mixture density estimation with historical floods and palaeoflood information, *ESAIM: PS*, July, Vol. 13, pg. 301-327.

Liang, K. Y. (1983). On information and ancilarity in the presence of a nuisance parameter, *Biometrika*, 70, pg. 607-612.

Lin, T. , Lee, J. and Yen, S. (2007). Finite mixture modelling using the Skew normal distribution, *Statistica Sinica,* 17, pg. 909-927.

Lindsay*,* B. G. (1995)*. Mixture models: Theory, geometry, and applications*. Hayward: Institute of Mathematical Statistics.

Liu, J. S. (2002). Mixture Kalman filters, *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), Vol 62, Issue 3, pg. 493–508.

Lo, Y., Mendall, N. R. and Rubin D. B. (2001). Testing the number of components in normal mixture, Biometrika, 88, pg. 767-778.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistics for the number of components in a normal mixture, *Applied Statistics*, 36, pg. 318-324.

McLachlan, G. J. and Peel, D. (1998). Robust cluster analysis via mixtures of multivariate t-distributions, *In Lecture Notes in Computer Science Vol. 1451, A. Amin, D. Dori, P. Pudil, and H. Freeman (Eds.). Berlin: Springer-Verlag,* pg. 658 - 666.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering.* New York: Marcel Dekker.

McLachlan, G. J. and Krishnan, T. (1997). The EM Algorithm and Extensions. *New York: Wiley.*

McLachlan, G. J. and Peel, D. (1996). An algorithm for unsupervised learning via normal mixture models, *In ISIS: Information, Statistics and Induction in Science, D.L. Dowe, K.B. Korb, and J.J. Oliver (Eds.*), pg. 354-363.

McLachlan, G. J., Peel D., Adams, P., and Basford, K. E. (1995). An algorithm for assessing by resampling the P-value of the likelihood ratio test on the number of components in normal mixture models. *Research Report No. 31. Brisbane: Centre for Statistics, The University of Queensland.*

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for reliability data.,* John Wiley & Sons, New York.

Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a degenerate rate and exponential distribution, *J. Indian Stat. Assn.*, Vol. 37, issue 2.

Muralidharan, K. (2000). The UMVUE and Bayes estimate of reliability of mixed failure time distribution, *Comm.Statist- Simulations & computations*, 29(2), 603-619.

Muralidharan, K. (2005). Estimation in presence of early failures, *Statistical Methods,* Vol. 7(2), 81-95.

Muralidharan K. and Arti M. (2008). Analysis of instantaneous and early failures in Pareto distribution, *Journal of statistical theory and Applications,* Vol 7, pg. 187-204.

Muralidharan K. and Arti M. (2011). A modified Pareto distrtibution, *Journal of the Indian Statistical Assoiation,* To appear in Vol. 49, No.1.

Muralidharan, K. and Kale, B.K. (2002). Modified Gamma distribution with Singularity at Zero, *Comm Statist- Stimulations & computations*, 31(1), pg. 143-158.

Muralidharan, K. and Kale, B. K. (2008). Inlier detection using Schawartz information criterion, *Journal. Reliability and Stat. Studies,* 1(1), pg. 1-5.

Muralidharan, K. and Lathika, P. (2004). The concept of inliers, *Proceedings of the First Sino-International Symposium on Probability, Statistics and Quantitative Management., Taiwan*, October, pg. 77-92.

Muralidharan, K. and Lathika, P. (2006). Analysis of instantaneous and early failures in Weibull distribution, *Metrika,* Vol 64, No.3, pg. 305-316.

Muralidharan, K. (2010). Inlier prone models: A review, *Probability Statistics Forum*, Vol. 3, pg. 38–51.

O'Neill, M. E. (1978). Distributional Expansions for Canonical Correlations from Contingency Tables, *Journal of the Royal Statistical Society, Series B (Methodological),* Vol 40, No. 3, pg. 303-312.

Pearson, K. (1894). Contributions to the Mathematical Theory of Evolution *Philosophical Transactions of the Royal Society A*, 185, pg. 71-110.

Peel, D. and Mmclachlan, G. (2000). Robust mixture modelling using the t distribution, *Statistics and Computing* 10**,** pg. 339–348.

Richardson, S. and Green, P. J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components, *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* Vol 59, Issue 4, pg. 731-792.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier detection*, John Wiley and Sons Ltd.

Schwarz, G. E. (1978). Estimating the dimension of a model, *Annals of Statistics,* **6** (2), pg. 461–464.

Singh, A. K. (2007). *Some contribution uniformly minimum unbiased estimations in exponential and irregular families of distributions,* P.hd thesis awarded by S.P.University, Vidyanagar.

Srivastava, M. S. and Worseley, K. J.(1986). Likelihood Ratio tests for a change in the multivariate normal mean, *Journal of the American Statistical Association,* 81, pg. 199-204.

Stephens, M. (2000). Dealing with label-switching in mixture models, *Journal of the Royal Statistical Society, Series B,* 62, pg. 795-809.

Sun, J. (1993). Tail Probabilities of the Maxima of Gaussian Random Fields, *The Annals of Probability*, Vol. 21, pg. 34-71.

Tan, W.Y. and Chang, W.C. (1972). Convolution approach to the genetic analysis of quantitative characters of self-fertilized population, *Biometrics,* 28: pg. 1073-1090 .

Titterington, D. M., Smith, A., Makov, U. E. (1985). *Statistical Analysis of finite mixture distribution,* John Wiley and Sons.

Vannman, K. (1991). Comparing samples from nonstandard mixtures of distributions with Applications to quality comparison of wood. Research report 1991:2 submitted to *Division of Quality Technology, Lulea University, Lulea, Swedon*.

Vannman, K. (1995). On the distribution of the estimated mean from the nonstandard Mixtures Of distribution, *Comm. Statist. – Theory and Methods*, 24(6), pg. 1569-1584.