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).

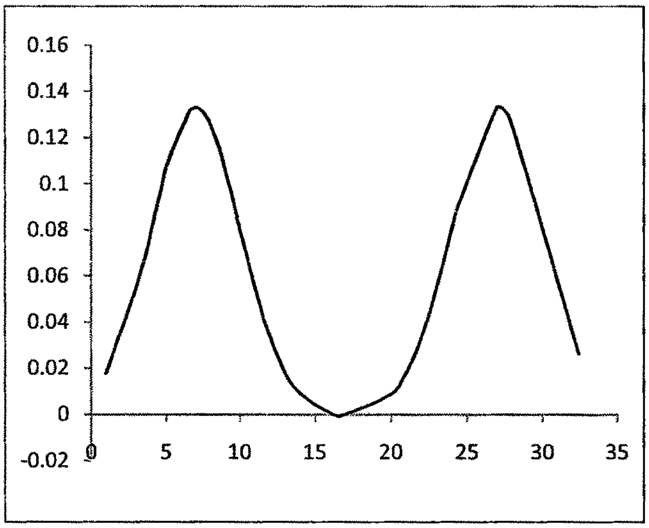

Fig. 5.3.1. Density function of mixture distribution

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

c.

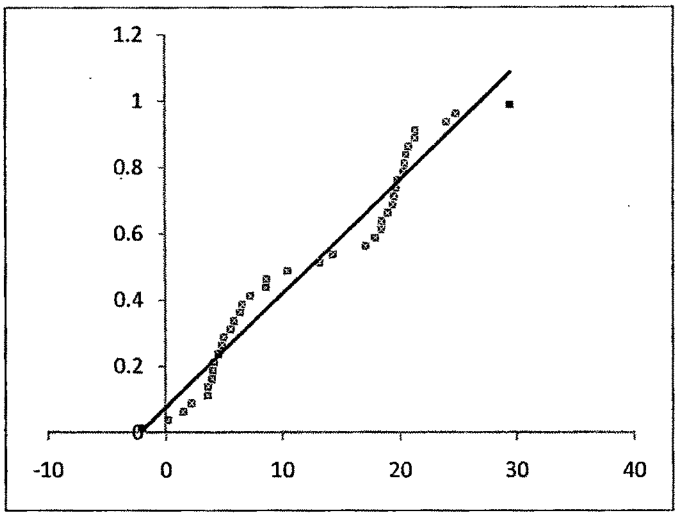

Fig. 5.3.2. Quantile-Quantile plot

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

$$\mu(\hat{\xi})=m \tag{5.3.1}$$

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

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

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 $\left\{\xi^{(m)}\right\}$ 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 $X = (x_1, x_2, \ldots.. x_n)$ be a sample of *n* independent observations from mixture of two inlier and target normal distributions and let $Z = (z_1, z_2, \ldots.. z_n)$ 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.

E- step

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

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

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

M-step

The quadratic form $Q\left(\theta\,|\,\theta^{(t)}\right)$ means that determining the maximizing values of θ is relatively straight forward. Firstly note that $p,(\mu_1,\sigma_1)\&(\mu_2,\sigma_2)$ 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

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

Then LMP test critical region is given by

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

where C is such that

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

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 H₀ if

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

where C′ is such that

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

5.4.2 Large sample test

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

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

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.

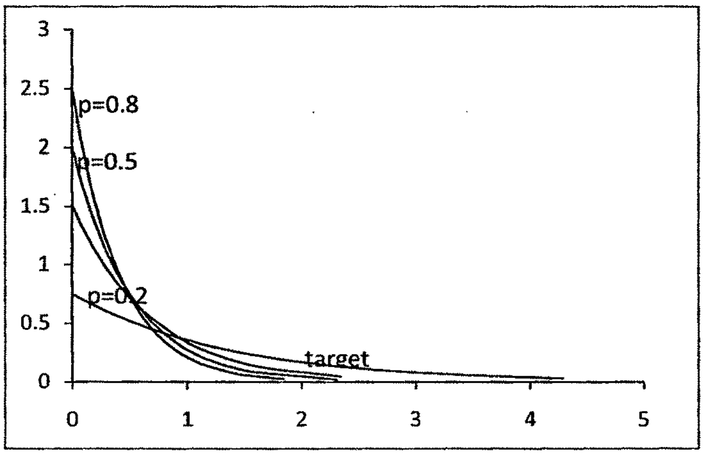

Fig. 5.5.1. Density function of exponential inliers and target distribution

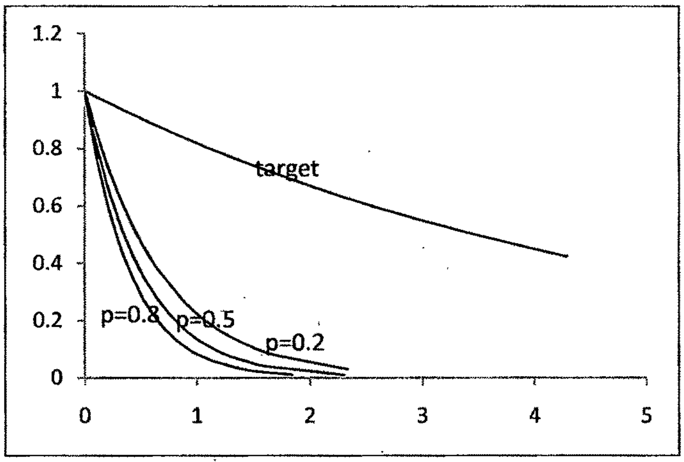

Fig. 5.5.2. Survival function of inliers and target in exponential distribution

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).

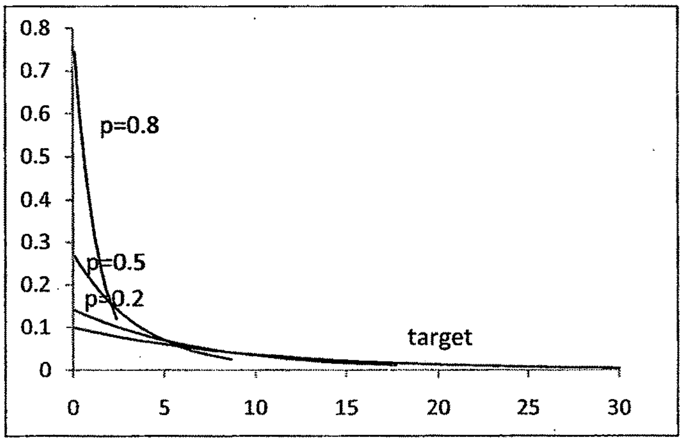

Fig. 5.5.3. Density function of exponential inliers and target distribution

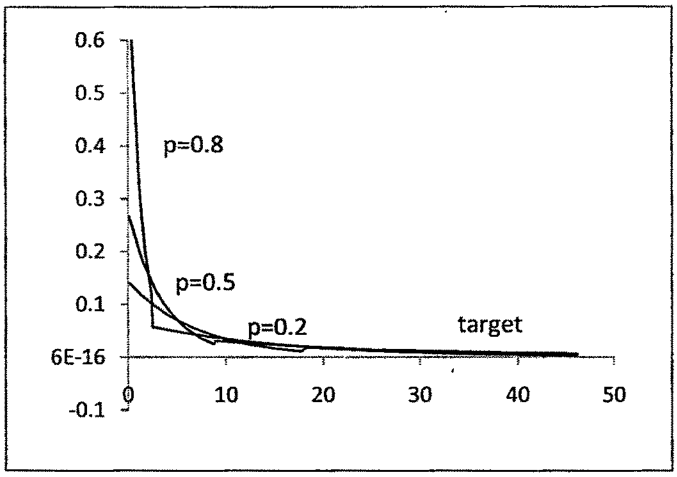

Fig. 5.5.4. Survival function of exponential inliers and target distribution

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)

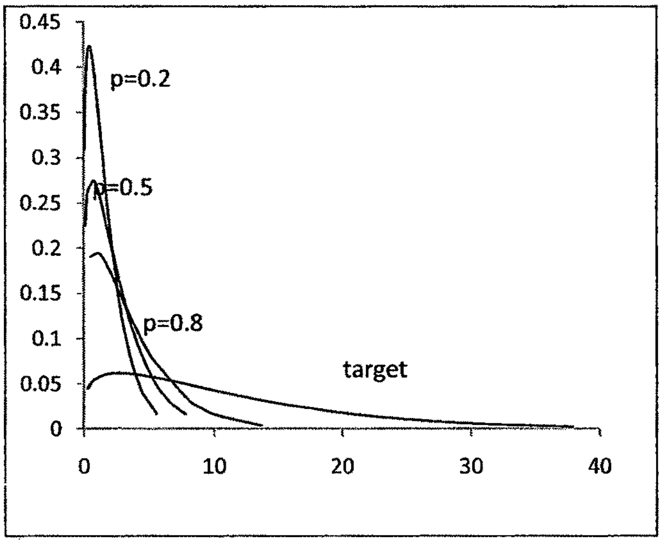

Fig. 5.5.5. Density function of Weibull inliers and target distribution

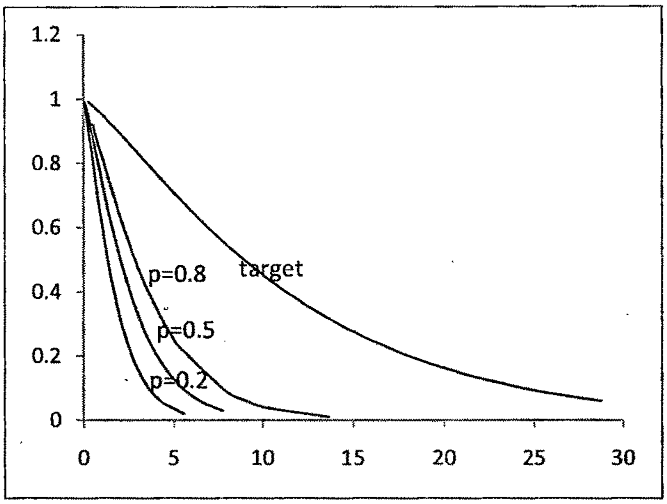

Fig. 5.5.6. Survival function of Weibull inliers and target distribution

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.

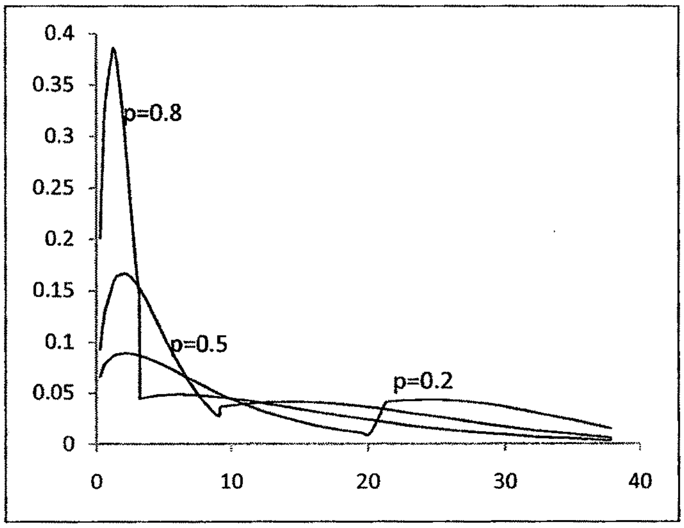

Fig. 5.5.7. Density function of Weibull inliers and target distribution

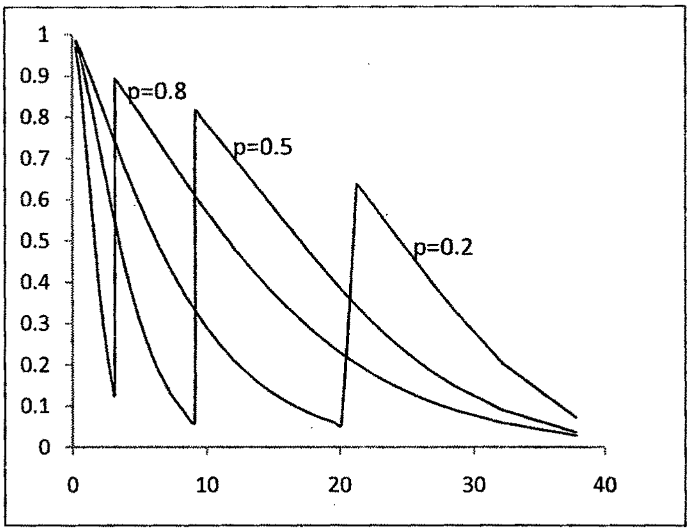

Fig. 5.5.8. Survival function of Weibull inliers and target distribution

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.

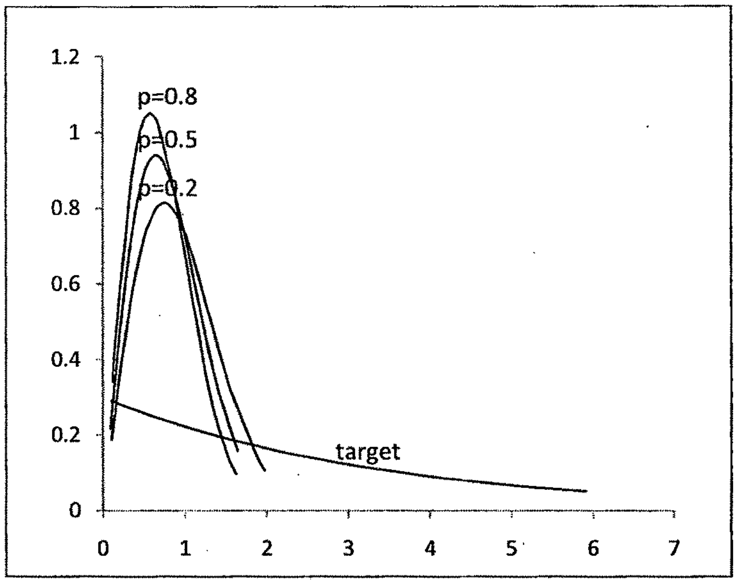

Fig. 5.5.9. Density function of Rayleigh inliers and exponential target population

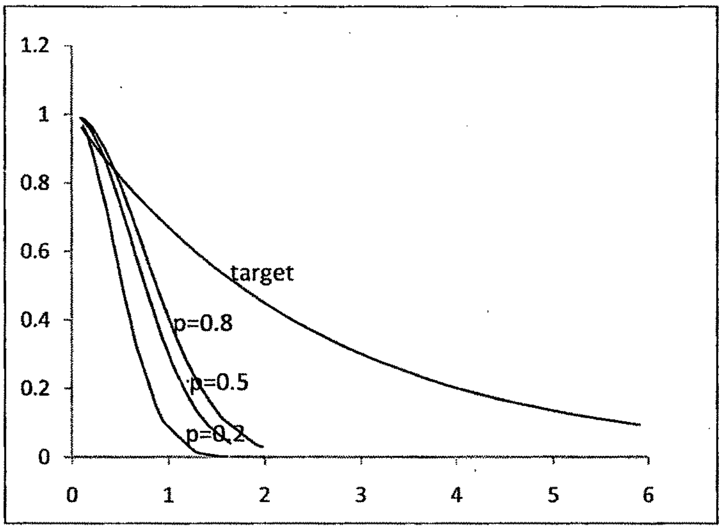

Fig. 5.5.10. Survival function of Rayleigh inliers and exponential target population

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)

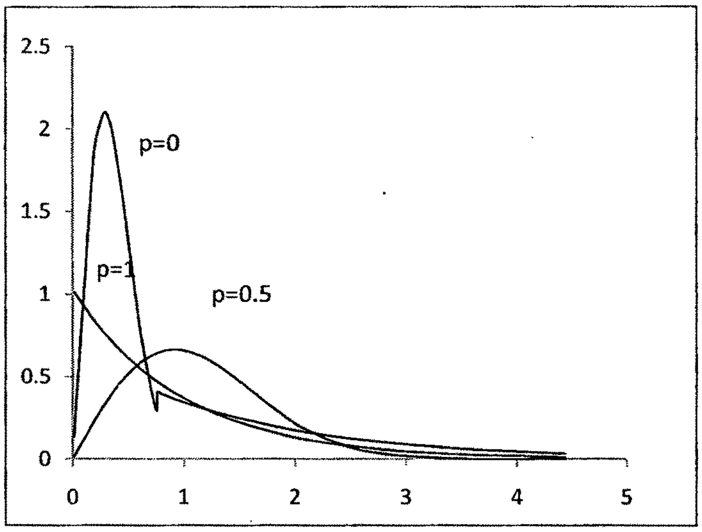

Fig. 5.5.11. Density function of Rayleigh inliers and exponential target population

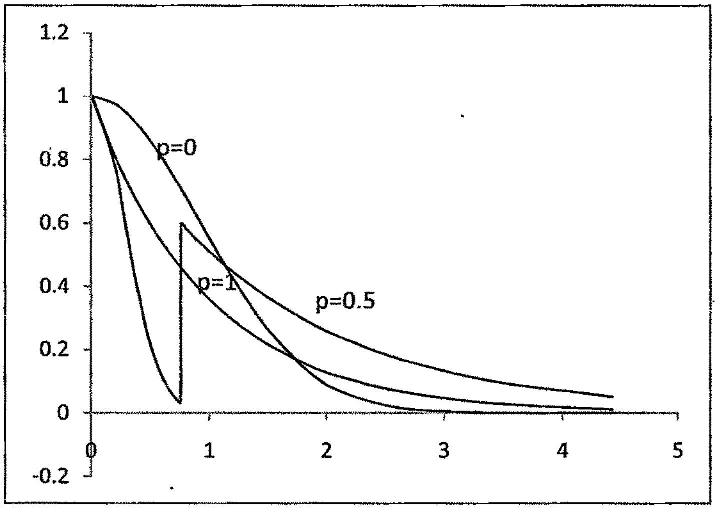

Fig. 5.5.12. Survival function of Rayleigh inliers and exponential target

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.